QC
1
F331
v. 18

22-A

AMERICAN JOURNAL of PHYSICS

A Journal Devoted to the Instructional and Cultural Aspects of Physical Science

VOLUME 18, NUMBER 1 JANUARY, 1950

The Teaching of Electricity and Magnetism at the College Level

I. Logical Standards and Critical Issues

(Report of the Coulomb's Law Committee of the A. A. P. T.)

TABLE OF CONTENTS

Part 1. Introduction

1.1 Scope and Plan of the Report
1.2 General Discussion of Logical Standards

Part 2. Critical Issues in the Logical Development of Electromagnetic Theory

2.1 Improper Use of Coulomb's Law in Electrostatics
2.2 Improper Use of Coulomb's Law in Magnetostatics
2.3 Polarization as a Basic Concept
2.4 Definitions of the Electric Vectors at Points Inside Ponderable Media
2.5 Concerning the Definitions of the Magnetic Vectors and Some Related Matters
2.6 Amperian Currents *versus* Magnetic Poles
2.7 Magnetic Fields in the Classroom

Part 3. Acceptable Schemes for Teaching the Fundamental Principles of Electricity and Magnetism

3.1 Introduction
3.2 Outline I: The Traditional Approach
3.21 The Electrostatic Field in Free Space
3.22 The Electrostatic Field in Dielectric Media
3.23 The Magnetostatic Field in Free Space
3.24 Electric Currents
3.25 Magnetic Materials
3.3 Outline II: The Amperian Approach
3.31 The Electrostatic Field in Free Space
3.32 The Measurement of Currents
3.33 D. C. Circuits

3.34 The Magnetic Fields of Steady Currents
3.35 Polarizable Media
3.36 Mechanical Forces with Iron
3.37 Fields in Iron

Appendices

Appendix A: Derivation of Formulas for the Total Force and Torque of Electrical Origin acting on a Charged Body Immersed in a Fluid Dielectric and Subject to an External Electric Field
Appendix B: The Behavior of Permanent Magnets in a Magnetic Fluid

List of Principal Symbols

F_e, $\mathbf{F}_e$ = net mechanical force of electrostatic origin including contribution from electrostriction.
f_e = direct electrostatic force of one charge on another.
q = electric charge (in formulating Coulomb's law).
q_c = conduction charge.
ρ_c, σ_c = volume and surface densities, respectively, of conduction charge.
ρ_p, σ_p = volume and surface densities, respectively, of polarization charge.
$\rho_t = \rho_c + \rho_p$ = volume density of *net* charge.
V = electric potential.
$\mathbf{P}$ = electric polarization.
κ_e = relative dielectric constant.
ϵ_0 = dimensional constant for D/E in free space.
$\epsilon = \epsilon_0 \kappa_e$.
χ_e = electric susceptibility (dimensionless).
I = conduction current.
$\mathbf{J}_c$, $\mathbf{J}_m$ = current densities for conduction currents and uncanceled Amperian currents, respectively.
$\mathbf{F}_m$ = net mechanical force of magnetic origin including contribution from magnetostriction.
$\mathbf{L}_m$ = net mechanical torque of magnetic origin.
p = pole strength (in formulating Coulomb's law for poles).
p_0, p_t = pole strength due to hard polarization only and to total polarization, respectively.
ρ_{m0}, ρ_m = pole density due to hard polarization only and to total polarization, respectively.
$\mathbf{M}_0$, $\mathbf{M}$ = magnetic polarization, hard and total, respectively.
μ_0 = dimensional constant for B/H in free space.
κ_m = relative permeability.
$\mu = \mu_0 \kappa_m$.
χ_m = magnetic susceptibility (dimensionless).
χ_m' = differential susceptibility of hard magnet.
$\eta = 1/\mu$ = reluctivity.
$\mathbf{n}$ = outward normal to surface.

1. Introduction

1.1 Scope and Plan of the Report

THE Coulomb's Law Committee was appointed in the fall of 1944 to formulate recommendations to the American Association of Physics Teachers based on a paper by Professor Carleton C. Murdock[1] criticizing the handling of Coulomb's law in some current elementary textbooks. The committee was further requested by President Lloyd W. Taylor to examine and report on the whole problem of the presentation of the fundamental concepts of electricity and magnetism in elementary courses in physics. It was suggested that it call attention to commonly occurring textbook errors and indicate ways of correcting them without making the texts unduly difficult.

The members of the committee approached their task from widely different points of view and with decidedly different biases. We have debated the issues at length. Certain crucial points have been subjected to detailed independent analyses by various members of the committee. We are pleased to report that on the basis of these investigations it has been possible to reach complete agreement as to the facts at issue. Regarding questions of taste and judgment our differences have been substantially reduced by protracted discussions. Where differences remain, the report presents both sides. In its present form it has the united support of all members of the committee.[2]

In attempting to carry out its assignment the committee has undertaken three major tasks,

[1] "Coulomb's law and the dielectric constant," *Am. J. Physics* **12**, 201–3 (1944).

[2] The constitution of the committee signing this report is as follows:

W. F. BROWN, JR., W. H. MICHENER,
N. H. FRANK, C. C. MURDOCK,
E. C. KEMBLE, *Chairman*, D. L. WEBSTER.

This list differs from the original one due to the regrettable resignation of Professor Norman E. Gilbert in 1946 and to the subsequent addition to the committee of Professors Murdock and Webster.

viz., (a) a careful examination of the facts of the theoretical situation with respect to the basic concepts of the classical theory of electricity and magnetism, (b) an examination of the practices commonly used in developing the basic concepts in current elementary textbooks of physics, and (c) a study of various possible correct methods of presenting the essential points of the theory to elementary and intermediate classes.

The report itself is arranged in three parts. Of these, *Part 1* is primarily devoted to an introductory discussion of the logical problems involved in teaching elementary physics on the college level. Here the reader will find a formulation of our general philosophy and also a resumé of our impressions regarding current logical standards in the writing of elementary college textbooks of physics. We did not find it necessary or feasible to make a collection of horrible examples of textbook errors, but have preferred to stress positive recommendations of acceptable practices.

Part 2 contains a detailed discussion of various critical issues in the logical development of electromagnetic theory. These include the original question of the proper application of Coulomb's law in electrostatics and magnetostatics, the problem of defining the various electric and magnetic vectors at points inside ponderable media, and finally a discussion of the roles to be assigned to magnetic poles and to Amperian currents in developing the theory of magnetism. In order to justify our conclusions as they affect elementary instruction, it has been necessary in Secs. 2.1 and 2.2 to delve into questions that go well beyond the scope of the elementary course. We believe, however, that the reader who wishes to do so will be able to skim much of the material in these sections without missing the main points.

Part 3 contains our positive recommendations in the form of two acceptable outlines, or synopses, for teaching the fundamental principles of electricity and magnetism, one using the traditional approach to magnetism by way of magnetic pole concept and the other using the new approach in which all magnetic fields are referred directly to electric currents of one kind or another. These outlines show how, in practice, we think the problem of logical standards might be handled in the teaching of electricity and magnetism. We hope the suggestions in Part 3 will be helpful to many teachers, but realize fully that there is nothing unique about our proposals.

Appendix A contains a derivation of fundamental importance for the discussion of forces on an electrified body in a fluid dielectric. *Appendix B* deals with the problem of relating the concept of the permanent magnet to the properties of actual ferromagnetic materials and especially with the effect of immersion in a magnetic fluid on the pole strength of a permanent magnet.

1.2 General Discussion of Logical Standards

From the point of view of the committee, the situation revealed by the study of current textbooks is rather discouraging. Although many authors have made a commendable effort to maintain high logical standards, the attempt to bring a complex and difficult theory down to the level of the poorly trained freshmen who throng so many of our colleges has too often led to confusion and absurdities. A full treatment of the theory involves the use of mathematical tools far beyond the reach of the college freshman and frequently unfamiliar to the authors of the textbooks themselves. Moreover, many of the students are unaccustomed to the precise use of abstract words and are unprepared to appreciate exact thinking. The problem of presenting the subject matter to elementary classes is, therefore, one that of necessity calls for some compromises between the requirements of logical rigor and precision, on the one hand, and the need for simplification as an aid to intelligibility, interest, and brevity, on the other. It is, perhaps, not surprising that the tendency has been to develop the ideas by successive approximations without great regard for logical rigor, treating special cases as if they were general, and relying in effect on the imitative solution of certain standardized classes of problems to act as a corrective to inaccurate statements in the text.

The situation seems to have been further complicated by the recent introduction of mks units, which, valuable as they may ultimately prove when fully standardized and universally accepted, involve a redefinition of some physical quantities and wholesale revision of basic formulas. The normal procedure for the author of an elementary textbook to use in checking the accuracy of his definitions is to refer to an advanced treatise of established merit. This procedure

has been made difficult by the abandonment of the conventions of the older books based on the Gaussian units. Two of the more prominent recent advanced texts[3] in electricity and magnetism differ not only from the older books as a result of adopting rationalized mks units, but differ from each other in fundamental formulas because they adopt different definitions of magnetic moment. Important as it is that the language of physics be expressed in terms of a wise choice of conventions, one cannot help wondering whether the difference between one choice and another is sufficient to justify an indefinite period in which a multiplicity of conventions operate in competition. We therefore urge that every effort be made to terminate the present transition period as soon as possible.

Recognition of the difficulties that beset the elementary teacher should not, however, lead to the easy acceptance of current practices and confusions. At least, such is the opinion of this committee. We believe that great improvement in the logical form of our elementary textbooks should be possible without making them unduly formidable. In our opinion the difficult points should be faced rather than side-stepped. At the top of the list of desiderata for a textbook should be the requirement that every statement be accurate and meaningful. Physics teachers betray their trust if they accept the logical standards which students bring with them to the classroom.

In this connection it may be permissible for us to emphasize the importance for physics teachers of the present widespread interest in semantics. The laws of physics are all expressed in a theoretical language involving concepts that we have created, or invented, in order to explain facts of experience. The laws are meaningless unless the conceptual schemes which lie behind them are properly understood. Moreover, these conceptual schemes have their origin and justification solely in their relation to experimental realities. A crucial test of good physics teaching is then the grasp which the student acquires of the relation between the word and letter symbols from which the language of physics is built and the experiments which they interpret. To give good instruction in this sense is the semantic problem of the physics teacher. If the student conquers the semantic difficulties of his subject, he can be expected to grow in self-confidence and in ability to do creative thinking in the field. If he fails on the semantic level, the subject matter becomes a mumbo-jumbo of words and verbal forms which he vainly attempts to memorize in the conviction that understanding is beyond him. Surely the last state of the student who is exposed to a course in physics, but fails to make the essential connections between symbols and experience, is worse than the first. Success with respect to some items, failure with respect to others, is, of course, the rule.

The discussion of the meanings of the electric vectors **E** and **D** at points inside solid or fluid materials given in Sec. 2.4 is an attempt to straighten out what is essentially a semantic problem.

While many statements in textbooks are undoubtedly wrong because the authors themselves have an incorrect understanding of the facts involved, others are at fault because the authors have chosen to introduce a general subject by the consideration of a special case and have not made clear to the student that the preliminary discussion is of limited scope. In some textbooks the general case is not mentioned. In others the more general case appears later without adequate reference to the previous treatment of a special case. For example, an author may define the permeability μ as a *constant* occurring in Coulomb's law and later disclose without explanation that in some materials μ is *not constant.*

The committee suggests that *authors should take pains to make clear at all times any limitations on the generality of the treatment of special topics.* One possible way to do this would be to supplement each equation by an italicized statement of the conditions under which it is applicable. This statement should preferably be separated from the rest of the text.

Carelessly worded definitions constitute a serious weakness of many elementary textbooks. In general, the definitions of physical quantities should not be wholly arbitrary. Ordinarily the definition of a new physical quantity is based on an experimental situation into which the quantity must fit. *Some reference to this experimental situation should accompany the definition.* It is confusing, for example, to define the magnetic flux through a coil of wire as that quantity whose time-rate of change is equal to the induced emf

[3] J. A. Stratton, *Electromagnetic theory* (McGraw-Hill, 1941) and G. P. Harnwell, *Principles of electricity and electromagnetism* (McGraw-Hill, 1938). The *Principles of electricity and electromagnetism* was issued in a second edition in 1949.

divided by the number of turns of wire, without making clear to the student that there is ample experimental evidence that the quantity so defined has an existence independent of the emf and can be computed, at least approximately, from the cross-sectional area of the coil, the nature and temperature of the material which constitutes the core, the distribution of magnetic intensity inside the core, and the previous magnetic history of the core. Unless the background physical situation is explained, the definition hangs in the air. One could arbitrarily define a physical quantity X to be associated with a body B as a quantity whose time-rate of change is equal to the volume of B, but such a definition would have no point, for X would not be associated with any significant physical law.

A satisfactory definition of a new physical quantity should indicate how the quantity is to be measured, or, if it is not directly measurable, how the term representing it is to be used.[4,5] Otherwise there is a logical gap before the definition can be put to work. Definitions by analogy, such as "Potential is electric pressure," do not meet this requirement and are to be avoided.

Finally, the definition of the unit for a physical quantity is not a satisfactory substitute for a definition of the physical quantity itself. Simply to say, for example, that the ampere is the current which, when passed through a single-turn circular coil of 1-cm radius, produces a force of $2\pi/10$ dynes on a unit magnet pole at its center, is not to give adequate instructions for the measurement of an arbitrary current.

Unfortunately, the setting up of definitions against their experimental background with complete logical rigor frequently leads to wordy statements boring to the average elementary student. Consider the definition of charge in electrostatics and its relation to Coulomb's law. It is customary to define the unit of charge and then to state Coulomb's law for point charges in a vacuum with no more ado. The law then gives an implicit completion of the definition of charge so that the student learns to use the concept correctly in the solution of problems, but the failure of the instructor to give a full and explicit definition works against the development of the habit of clear thinking in the mind of the student.

The following discussion may serve to represent the result of an attempt on the part of the ultraconscientious instructor to tell the story with complete logical rigor and clarity. Even to an experienced teacher of electrical theory it may be surprising to see how long the story is.

After exhibiting the elementary phenomena of frictional electricity and explaining the usual convention regarding the signs of charges, he confronts the class with the need for devising a workable scheme for the measurement of electric charges. He notes that any scheme of this kind within the framework of electrostatics must be based upon the forces which charges exert on one another. He observes that air, itself, under the influence of charged bodies, exerts on them forces which are quite small in comparison with the forces they exert on one another. Ideally the experiments to be considered should all be performed in a vacuum, but if they are actually performed in air and analyzed as if performed in a vacuum, the error will be small. On the basis of this introduction he proceeds to develop the experimental facts and conventions somewhat as follows:

(*a*) *Experimental fact.*—The force between two charged bodies is constant in time if the bodies are well insulated and are kept at a fixed distance apart.

(*b*) *Convention.*—The charges on a set of bodies are to be reckoned as constant so long as the forces of interaction depend only on the relative positions of the bodies.

(*c*) *Experimental fact.*—The force between two bodies carrying constant charges depends on their distance of separation r, on their shape, and on the way in which the charges are distributed over the bodies. However, if the linear dimensions of the bodies are small compared with r, the shape and the distribution are immaterial. In this case the force varies as $1/r^2$.[6]

(*d*) *Convention and corollaries.*—It is convenient to introduce the term "point charge" for an idealized charge concentrated at a geometric point. The force between point charges must then be supposed to vary as $1/r^2$ at all distances. A charge spread over a body of negligible dimensions approximates a point charge and can be treated like a point charge. Any distribu-

[4] See discussion in Sec. 2.4.

[5] Sometimes it is pedagogically helpful to supplement the definition of a new concept with information regarding interpretations not contained in the definition. Thus the definition of charge can be accompanied by the explanation that the quantity defined can ultimately be interpreted as a number of particles and the definition of magnetic moment may be supplemented by observing that a current loop has a moment which is equal to, or proportional to, the product of current and area.

[6] The instructor should note at this point that there are indirect methods of demonstrating the inverse-square law far more accurate than direct measurements of the forces on small charged bodies.

tion of charge can presumably be regarded as equivalent to a distribution of point charges.

(*e*) *Convention.*—The charges on two bodies shall be said to be *equal* in magnitude if they exert equal forces on a third charge equidistant from them, the distance being such that all three can be treated as point charges.

(*f*) *Experimental fact.*—If a charge X is subject to simultaneous electric forces from two or more charges $A, B, C, \cdots$, the net force which it experiences is the vector sum of the forces which it would experience if the several charges $A, B, C, \cdots$, were to act one at a time.

(*g*) *Corollary.*—If point charges $A, B, \cdots$, are brought together at a point P, the force which the combination of charges exerts on a point charge X is numerically equal to the algebraic sum of the force components along the line PX which they would separately exert if placed one after the other at P.

(*h*) *Corollary.*—If n equal point charges of the same sign are brought together at the point P, the force that the combination exerts on any other point charge X will be n times the force which any one of them would exert if acting by itself.

(*i*) *Definition.*—The ratio of the numerical values of any two point charges is to be identified with the ratio of the forces which they individually exert on a third point charge at the same distance r.

(*j*) *Definition.*—Let the unit point charge be defined (provisionally) as one which exerts a force of magnitude k_0/r^2 on an equal point charge at a distance of separation r. The numerical value to be assigned to k_0 varies from one system of units to another in a manner to be described hereafter.

The definitions of the unit point charge and of the ratio of the numerical values of any two point charges serve to define the numerical value of any point charge. The algebraic value q is obtained by multiplying the numerical value by ± 1 according to the usual sign convention. The value of an extended charge is defined by the same rule of measurement with the restriction that r must always be large compared with the linear dimensions of the bodies on which the charges are located.

(*k*) *Corollary.*—The total charge created by bringing together a number of separate charges is equal to the sum of their algebraic values expressed in terms of a common unit.

(*l*) *Corollary on* (*i*).—The force exerted by a point charge of magnitude $|q|$ on any other point charge X at a distance r is $|q|$ times the force exerted by a unit charge on X at the same distance.

(*m*) *Corollary on* (*c*) *and* (*l*).—The magnitude of the force exerted on each other by any two point charges of magnitude $|q|$, $|q'|$, respectively, whose distance apart is r, is

$$|F| = k_0 |q|\,|q'|/r^2. \qquad (1\text{–}1)$$

If the algebraic value of the force be defined as positive for repulsion and negative for attraction, this equation can be replaced by

$$F = k_0 qq'/r^2. \qquad (1\text{–}2)$$

This is Coulomb's law of force for point charges.

(*n*) *Corollary.*—The law (1–2) holds for extended charges provided that r is large compared with their maximum linear dimensions.

The instructor then proceeds with the specification of the value of k_0 in the system or systems of units which he chooses to employ.

In conclusion the instructor should point out that, although this outline gives a useful logical analysis of the way in which the theory is constructed from elementary observations, it does not describe the intuitive mental processes by which theories are ordinarily created; nor does it give an adequate conception of the wealth of evidence for the theory contained in the wide range of experimental tests that have been made to check the total consequences as well as the basic assumptions.

This method of presenting the first steps in electrostatics is admittedly involved and tedious. A teacher of elementary physics who desires to maintain a high logical level while holding the interest of his class is thereby confronted with a delicate pedagogical problem. Fortunately, his story can be made less formidable by introducing an operational definition of charge based on the Faraday ice-pail experiment and independent of direct reference to Coulomb's law. Or the presentation can be simplified by a quasi-deductive postulational approach. These permissible modifications of procedure are described in Sec. 3.21 of Part 3. Suffice it to say here that every teacher and textbook writer should realize that there are real hazards to clear thinking in a very radical cutting of the argument. Physics majors should at some time be exposed to the full story. Even in freshman courses complete failure to define the ratio of two charges is not to be recommended.

2. Critical Issues in the Logical Development of Electromagnetic Theory

2.1 Improper Use of Coulomb's Law in Electrostatics

The first issue confronting the committee relates to the use of Coulomb's law in the form[7,8]

[7] To avoid misunderstanding, the equations in this report will ordinarily be written in both Gaussian and mksc units. The Gaussian system is the much-used mixed scheme in which all electrical quantities are expressed in unrationalized cgs electrostatic units, whereas the magnetic

(Gauss) $\quad$ "$F_e = qq'/\kappa_e r^2$," $\quad$ (2–1a)
or (Giorgi) $\quad$ "$F_e = qq'/4\pi\epsilon_0\kappa_e r^2$," $\quad$ (2–1b)

and the equivalent magnetic Eqs. (2–15) as the respective basic equations of the electrostatics of dielectric materials and the magnetostatics of magnetic materials. Although the corresponding equations for free space, Eqs. (2–2), can properly occupy a central position in the development of electrostatics and magnetostatics, it required little deliberation to disclose that the committee was from the beginning unanimously of the opinion that Eq. (2–1) and its magnetic analog Eq. (2–15) are not of general validity and are, therefore, not suitable for use as the basic equations of electrostatics and magnetostatics. In this the committee supports the primary thesis of the previously mentioned article by Murdock.[1] It seemed desirable, nevertheless, to make a critical examination of these two much-used, but dubious, equations. The conclusions to which we have been led are reviewed herewith, first for the electrostatic case and then, in Sec. 2.2, for the magnetostatic case.

Since the classical theories of the electrostatics and magnetostatics of material media have a much higher standing than any available experiments designed to give a direct check on the validity of the equations under consideration, our discussion must take a theoretical form.

The reader will recollect that the classical theory of electrostatics, in harmony with the Lorentz electron theory, postulates that Coulomb's law in the primitive form

(Gauss) $\quad f_e = qq'/r^2$, $\quad$ (2–2a)
or (Giorgi) $\quad f_e = qq'/4\pi\epsilon_0 r^2$, $\quad$ (2–2b)

applies in a ponderable medium as well as in a vacuum, provided we identify f_e with the *direct* force of one charge on another. The apparent influence of a ponderable medium on the observable force F_e exerted by one charge on another is assumed to be due to the electrical structure of matter and to the distribution of electric polarization created in the medium by an electric field. The electric polarization $\mathbf{P}$ gives rise to polarization charges in addition to the conduction charges. The potential V and the electric intensity $\mathbf{E}$ are to be computed as in a vacuum, except that the polarization charge densities are included with the conduction charge densities to give total charge densities which are the sources of the field.

The observable force $\mathbf{F}_e$ acting on an element of conduction charge in the presence of dielectrics is the vector sum of the direct forces $\mathbf{f}_e$ exerted on it by other conduction charges, the direct forces exerted on it by the polarization charges, and a mechanical force of electrostrictive origin.

The relation between the polarization vector $\mathbf{P}$ and the electric intensity $\mathbf{E}$ depends on the nature of the medium. If $\mathbf{P}$ is a homogeneous linear function of $\mathbf{E}$, the medium is conveniently designated as *linear*. If it is nonconducting, we call it a *dielectric*. If it is homogeneous in the sense that its properties are independent of the positional coordinates, and if it is electrically isotropic, $\mathbf{P}$ is proportional to $\mathbf{E}$ and throughout the medium we have an electrical susceptibility and a dielectric constant with unique values. Most treatments of dielectric behavior apply only to such media. Clearly all these restrictions are necessary if Eq. (2–1) is to be meaningful. On the other hand, there is no fundamental difficulty in dealing with crystalline media which are electrically anisotropic, if a linear vector operator (dyadic) is used to relate $\mathbf{P}$ and $\mathbf{E}$. There are also media to which the linearity postulate, as formulated above, does not apply,

---

quantities are expressed in the unrationalized cgs electromagnetic units.

The second system employed is the rationalized Giorgian meter-kilogram-second-coulomb system. Since there are two different sets of proposals regarding the units of magnetic pole-strength, magnetic moment, and magnetization within the mksc scheme, and since the merits of these two systems come up for discussion in the report (see Sec. 2.6), we indicate where necessary whether an equation follows the Sommerfeld or the Kennelly definitions.

In general, the different forms of the same relationship in different unit-schemes are labeled with the same equation number, but with different letters.

Vector quantities are indicated by bold-faced type.

Quotation marks around Eqs. (2–1) and (2–15) indicate that these equations are introduced not as statements of fact, or of accepted theory, but as assertions whose validity is in question.

The reader is warned that on account of an error in the preparation of the manuscript there are no equations for the numbers (2–30), (2–31), and (2–32).

[8] The choice of symbols for a report that uses both Gaussian and Giorgian units is a bit difficult. To avoid ambiguities we denote the relative dielectric constant and the relative permeability by κ_e and κ_m, respectively. The basic mksc dimensional constants ϵ_0 and μ_0 represent the ratios $\mathbf{D}/\mathbf{E}$ and $\mathbf{B}/\mathbf{H}$ for free space. Although ϵ and μ are now commonly used for the products $\epsilon_0\kappa_e$ and $\mu_0\kappa_m$, respectively, it has seemed best for our present purposes to write out these products explicitly.

e.g., piezoelectric and pyroelectric crystals and materials which show saturation effects.

It follows from these considerations that *Eq. (2–1) is not applicable throughout the whole range of electrostatic problems and, on that account, cannot be considered fundamental.* The meaning and validity of the equation for problems involving only homogeneous, isotropic, and linear dielectrics—let us say *Class A dielectrics*—remains to be discussed.[9]

Since the point charges of Coulomb's law are ideal concepts, it is important to stipulate the physical interpretation of Eq. (2–1) in connection with a discussion of its validity. The simplest and most direct interpretation identifies F_e with the limit of the force of interaction between two small charged bodies immersed in a fluid dielectric when the linear dimensions of the bodies are made very small compared with their distance of separation. This may be designated the *point-charge interpretation.*

A second interpretation identifies F_e with the differential force between two charge elements in a continuous distribution of charge. The equation can then be written in vector form as[10]

$$\text{(Gauss)} \qquad d\mathbf{F}_e = \frac{dq_c dq_c'}{\kappa_e r^2}\mathbf{1}_r, \tag{2–3a}$$

$$\text{or (Giorgi)} \qquad d\mathbf{F}_e = \frac{dq_c dq_c'}{4\pi\epsilon_0 \kappa_e r^2}\mathbf{1}_r, \tag{2–3b}$$

where $\mathbf{1}_r$ is a unit vector directed from the source-charge dq_c' to the charge dq_c which experiences the force. This may be designated the *continuous charge interpretation.* Finally we have the possibility, needing no explicit discussion, of the forces of interaction between a point charge and a continuous distribution of charge.

Whether we have in mind point charges or continuous distributions it is convenient to resolve the problem of the validity of the equations in question into two subproblems by introducing the electric intensity $\mathbf{E}'$ due to q_c' as an intermediate factor. Equation (2–3) is then replaced by

$$\text{(Gauss)} \qquad d\mathbf{E}' = \frac{dq_c'}{\kappa_e r^2}\mathbf{1}_r, \tag{2–4a}$$

$$\text{or (Giorgi)} \qquad d\mathbf{E}' = \frac{dq_c'}{4\pi\epsilon_0 \kappa_e r^2}\mathbf{1}_r, \tag{2–4b}$$

$$\text{and} \qquad d\mathbf{F}_e = dq_c \mathbf{E}'. \tag{2–5}$$

Our first question is accordingly, "What is the standing of Eq. (2–4) for point charges q_c' and for continuous distributions of charge?" In answer, we first point out that, according to the basic definition of the theory,[11]

$$\text{(Gauss)} \qquad \mathbf{E} = \int \frac{\mathbf{1}_r dq_t}{r^2}, \tag{2–6a}$$

$$\text{or (Giorgi)} \qquad \mathbf{E} = \int \frac{\mathbf{1}_r dq_t}{4\pi\epsilon_0 r^2}. \tag{2–6b}$$

[9] In Maxwell's *Treatise on electricity and magnetism* (Oxford, 1904), 3rd ed., Coulomb's law is found only in the form (2–2) in passages written by Maxwell himself; moreover, both in the electrostatic statement of the law (Vol. 1, pp. 45–46) and in the magnetostatic (Vol. 2, p. 3), the original wording contains no mention of the medium. At both these points the editor, J. J. Thomson, felt it necessary to insert (in curly brackets; see Vol. 1, p. xvi) the qualification that the medium is supposed to be air; and on p. 122 of Vol. 1, at the conclusion of the discussion of systems of conductors in a liquid, he inserted the following sentence: "It follows from the preceding investigation that the force between two electrified bodies surrounded by a medium whose specific inductive capacity is K is $ee'/Kr^2\cdots$." It is clear that Maxwell himself did not consider Eq. (2–1) either basic or important and that "the force between two bodies" was to him an unambiguous concept, not requiring any mention of the intervening medium.

A reading of Maxwell's preface and introductory chapters makes it evident why this was so. Maxwell's object was the mathematical formulation of Faraday's concepts, which located all action in the medium, and the demonstration that when so formulated they were *mathematically equivalent* to the theory based on action at a distance. The two theories, however, are *conceptually distinct.* When one speaks of the force between two bodies, one is using the action-at-a-distance theory; this theory, in Maxwell's own words (Vol. 1, p. 155), "supposes the electric force to act directly between bodies at a distance, no attention being bestowed on the intervening medium." If one wishes to use the Faraday-Maxwell theory, one must—whether in a vacuum or in a material dielectric—speak not of the force between the bodies but of the force between adjacent portions of the stressed medium. The idea that the "force between the charges" can "depend on the medium" represents an incongruous mixture of the two theories, never used by Maxwell but introduced later by his followers. Equation (2–1) occurs in Heaviside's papers, as well as in the editorial insertion by Thomson mentioned above.

[10] From this point on we shall use subscripts c, p, and t, respectively, for conduction charges, polarization charges, and total charges.

[11] In Eqs. (2–4) and (2–5) the primes are used to distinguish between the external charges q_c' (producing the external field $\mathbf{E}'$) and the charges q_c on which the force $\mathbf{F}_e$ acts. In Eqs. (2–6)–(2–9) this distinction is dropped because the relations hold for any charges and corresponding field.

Here dq_t is an element of *total charge*, including polarization charge as well as conduction charge. The integral of Eq. (2–6) as written is intended to include contributions from volume densities, surface densities, line densities and, in the limit, point concentrations of charge.

Evidently the replacement of Eq. (2–6) by Eq. (2–4) is valid if, and only if, the charge elements dq_t are everywhere equal to the corresponding conduction charges dq_c divided by κ_e. But if, for simplicity, we confine our discussion for the moment to volume distributions of charge and stick to Gaussian units, the basic relation between the two kinds of charge[10] is

$$\text{(Gauss)}\quad \mathrm{div}\mathbf{E}/4\pi=\rho_t=\rho_c+\rho_p=\rho_c-\mathrm{div}\mathbf{P}. \qquad (2\text{–}7)$$

To justify the replacement we must first assume that we have to do with linear isotropic dielectrics. This special assumption yields

$$\left.\begin{aligned}&\text{(Gauss)}\\ &\quad \mathrm{div}\mathbf{E}/4\pi=\rho_c-\mathrm{div}(\chi_e\mathbf{E})\\ &\qquad\quad =\rho_c-\chi_e\,\mathrm{div}\mathbf{E}-\mathbf{E}\cdot\mathrm{grad}\chi_e,\end{aligned}\right\} \qquad (2\text{–}8)$$

and, since $1+4\pi\chi_e=\kappa_e$,

$$\text{(Gauss)}\qquad \rho_t\kappa_e=\rho_c-\mathbf{E}\cdot\mathrm{grad}\kappa_e. \qquad (2\text{–}9)$$

Here χ_e denotes the electrical susceptibility. Thus the validity of Eq. (2–4) requires not only the restriction of the discussion to linear isotropic dielectrics, but also that $\mathbf{E}\cdot\mathrm{grad}\kappa_e$ shall vanish throughout the field. If surfaces of discontinuity occur in the field the normal component of the intensity on these surfaces must vanish. *In practice this means that all parts of space in which there is a field must be filled with a single Class A dielectric if Eq. (2–4) is to be valid.* Conductors in electrostatic equilibrium may be present since the field inside them vanishes. Of course, Eq. (2–4) can be used to compute a useful approximation when the foregoing condition is not completely fulfilled provided that the field is sufficiently weak at the boundaries of the homogeneous dielectric and provided that the point at which the intensity is being computed is at a sufficient distance from such boundaries. The restrictions are very severe, however. *The equation is definitely not of general applicability if the dielectric is inhomogeneous, electrically anisotropic, or nonlinear.*

We turn now to Eq. (2–5) and consider first applications to extended charge distributions q_e on bodies immersed in Class A fluid dielectrics. The total force $\mathbf{F}_e$ acting on such a body X consists of two parts: the force due to *direct* electrical interaction with other charged bodies and with the fluid; and the resultant mechanical force acting on X due to the asymmetry of the distribution of mechanical pressure around it resulting from electrostriction in the fluid. Investigation of this force $\mathbf{F}_e$ may be carried out by energy methods or by direct examination of the force contributions from various sources. The usual procedure is to use the energy method to work out the body force and surface forces acting in an electrostatic field and responsible for electrostriction. These can then be integrated directly over the body or replaced by a stress system to compute the total force $\mathbf{F}_e$ to be balanced by mechanical supports. We shall concern ourselves here solely with this total force, ignoring the problem of its distribution over X.

If the vectors $\mathbf{D}$ and $\mathbf{E}$ have the same direction in the body X, so that we can use the relation $\mathbf{D}=\kappa_e\mathbf{E}$, and if we postulate the necessary transition layers to replace surface contributions by volume contributions, the total force $\mathbf{F}_e$ can be computed from the total electric intensity $\mathbf{E}$ by the equation[12]

$$\text{(Gauss)}\quad \mathbf{F}_e=\int_X\left[\rho_c\mathbf{E}-\frac{E^2}{8\pi}\mathrm{grad}\kappa_e\right]d\tau, \qquad (2\text{–}10)$$

where $d\tau$ is an element of volume. It is immediately evident from this equation that Eq. (2–5) cannot be of general validity. In fact, as is well known, an uncharged dielectric body whose dielectric constant differs from that of the surrounding fluid will experience a net force different from zero when placed in a nonuniform external electric field.

An alternative and, in some respects, more general equation for the total force acting on the body X is

$$\mathbf{F}_e=(\kappa_e)_A\int_X\mathbf{E}\rho_t d\tau, \qquad (2\text{–}11)$$

where $\mathbf{E}$ is again the total electric intensity, ρ_t is

[12] Cf. J. A. Stratton, Ref. 3, Secs. 2.26, 2.29, and especially Eq. (13), p. 158.

the total charge density including the polarization charge, and $(\kappa_e)_A$ is the relative dielectric constant of the Class A fluid which surrounds X. The integration is to be extended over all charges in and adjacent to the surface of X including polarization charges.[13]

There are three important special cases in which the general equation (2–11) can be reduced to the special form envisaged in Eq. (2–5) for an extended distribution of charge, namely,

$$\mathbf{F}_e = \int_X \mathbf{E}' \rho_e d\tau. \tag{2–12}$$

The first case is the one already discussed in which X has the same dielectric constant as the surrounding fluid. One qualification only is needed for this case in view of the fact that in Eq. (2–12) $\mathbf{E}'$ is the intensity due to the *external* distribution of conduction charge q_c'. The total intensity can be resolved into the sum of three terms as follows:

$$\mathbf{E} = \mathbf{E}_s + \mathbf{E}_0' + \Delta\mathbf{E}', \tag{2–13}$$

where $\mathbf{E}_s$ is the self-field of the total charge distribution on X, $\mathbf{E}_0'$ is the external field before the body X was placed in position, and $\Delta\mathbf{E}'$ is the field due to the alteration in the net charge distribution on other bodies due to the introduction of X and its charge. Since the self-field can exert no resultant force on X it is correct to substitute $\mathbf{E}_0' + \Delta\mathbf{E}'$ for $\mathbf{E}$ in Eq. (2–10) after eliminating the term involving grad κ_e. Hence we must understand *either* that $\mathbf{E}'$ in Eq. (2–12) means the external field *as modified* by the presence of X and its charges (i.e., $\mathbf{E}_0' + \Delta\mathbf{E}'$), *or else* that the size of X, its dielectric constant, and its charge distribution are all so chosen that $\Delta\mathbf{E}'$ is negligible.

The second case is the one in which the body X is a conductor carrying a surface density of conduction charge σ_c. The corresponding surface density of total charge is $\sigma_c/(\kappa_e)_A$. In this case Eq. (2–11) reduces at once to Eq. (2–12) or to the equivalent surface integral. The same remarks about the interpretation of $\mathbf{E}'$ apply in this case as in the previous one.

The third case is that in which $\mathbf{E}_0'$ is constant and $\Delta\mathbf{E}'$ is negligible. Replacing $\mathbf{E}$ in Eq. (2–11) by $\mathbf{E}_0'$ we obtain

$$\mathbf{F}_e = (\kappa_e)_A \mathbf{E}_0' \int_X \rho_t d\tau = (\kappa_e)_A \mathbf{E}_0' \int_S E_n dS$$

$$= \mathbf{E}_0' \int_S D_n dS = \mathbf{E}_0' q_c,$$

where dS is an element of surface. This completes our examination of the applicability of Eqs. (2–5) and (2–12) to cases in which the charge q_c is distributed.

We turn finally to the point-charge case which most teachers of elementary courses envisage in connection with Coulomb's law. It is not difficult to see in the light of Eq. (2–11) that if we allow the body X to shrink to zero size about some fixed point P, the charge in each volume element being carried over at each stage into an equal charge in the corresponding reduced volume element, then the force $\mathbf{F}_e$ will approach the value $q_c\mathbf{E}'$, where $\mathbf{E}'$ is the external field at the point P in the presence of the point charge q_c. This is because the multipole moments of the distribution on X approach zero as the linear dimensions of the body are reduced.[14] We conclude that Eq. (2–5) is valid for a small charged body immersed in a Class A fluid dielectric in the sense that

$$\mathbf{E}' = \lim_{a \to 0} (\mathbf{F}_e/q_c),$$

where a is a linear dimension of the body and the limiting process indicated is to be carried out in the manner just described. If $\mathbf{E}_0'$ is desired, it will suffice to let q_c as well as a approach zero. Then $\mathbf{E}'$ passes over into $\mathbf{E}_0'$, and we can write

$$\mathbf{E}_0' = \lim_{\substack{a \to 0 \\ q_c \to 0}} (\mathbf{F}_e/q_c). \tag{2–14}$$

Thus the theory tells us that $\mathbf{E}$ *should be equally measurable in a Class A dielectric fluid and in a*

[13] For a brief derivation of Eq. (2–11) and a proof of Eq. (2–10) based on Eq. (2–11), see Appendix A.

[14] Introducing $\bar{\rho}_t$ for the mean value of ρ_t on X and defining $\delta\rho_t$ by the relation $\delta\rho_t = \rho_t - \bar{\rho}_t$, we can replace Eq. (2–11) by

$$\mathbf{F}_e = (\kappa_e)_A \left[\bar{\rho}_t \int_X \mathbf{E}' d\tau + \int_X \mathbf{E}' \delta\rho_t d\tau \right].$$

If the linear dimensions of X are reduced to zero the first term will approach $q_c\mathbf{E}(P)$ and the second term, which involves only the multipole forces, will approach zero.

vacuum by the traditional force-per-unit-test-charge method.[15]

This conclusion is, of course, in harmony with the interpretation of **E** as the force-per-unit-charge acting on the movable charges in the theory of current electricity. It can be simply checked by means of a mental experiment if one postulates that the dielectric model is to be in harmony with the principle of conservation of energy. Imagine a fluid dielectric in a container with insulating side walls and charged conducting plates A and B at top and bottom. Let the space around the container be evacuated. Then it becomes possible in principle to carry a small test-charge slowly from A to B either through the fluid or through the evacuated space. For the energy to be conserved, the work done against the electrostatic forces must be the same in the two cases. Hence, the work done in the fluid as well as in the vacuum must equal the product of the charge and the potential difference through which it is carried.[16]

As a corollary on our last result and the previous discussion of the validity of Eq. (2–4), we infer that Eq. (2–1) holds for "point charges" in an infinite, Class A dielectric fluid. It would *not* hold for point charges near the interface of two dielectrics or near a conducting surface.

In view of the limitations on the validity and usefulness of Eq. (2–1) and the likelihood of misunderstanding regarding it, we believe that it would be better not to introduce it into courses in general physics for beginners.

2.2 Improper Use of Coulomb's Law in Magnetostatics

We turn next to the magnetic form of Coulomb's law analogous to Eq. (2–1) and given by

$$\text{(Gauss)} \qquad \text{“}F_m=\frac{pp'}{\kappa_m r^2}\text{”} \qquad (2\text{–}15a)$$

$$\text{(Giorgi-Kennelly)} \qquad \text{“}F_m=\frac{1}{4\pi\mu_0}\frac{pp'}{\kappa_m r^2}\text{”} \qquad (2\text{–}15b)$$

$$\text{(Giorgi-Sommerfeld)} \qquad \text{“}F_m=\frac{\mu_0}{4\pi}\frac{pp'}{\kappa_m r^2}\text{.”} \qquad (2\text{–}15c)$$

(Quotation marks indicate that the validity of this equation is in question. See footnote 7.)

The classical theory of magnetostatics starts with Coulomb's law in the form analogous to Eq. (2–2):

$$\text{(Gauss)} \qquad f_m=\frac{pp'}{r^2} \qquad (2\text{–}16a)$$

$$\text{(Giorgi-Kennelly)} \qquad f_m=\frac{1}{4\pi\mu_0}\frac{pp'}{r^2} \qquad (2\text{–}16b)$$

$$\text{(Giorgi-Sommerfeld)} \qquad f_m=\frac{\mu_0}{4\pi}\frac{pp'}{r^2}. \qquad (2\text{–}16c)$$

Although this equation usually is stated, as here, for point poles and is illustrated by Coulomb's experiments, the theory does not depend on the possibility of approximating point poles experimentally; and in precise analyses of experimental situations, the poles are supposed distributed over surfaces and throughout volumes. It is postulated that Eq. (2–16) gives the direct force f_m of one pole on another, in a ponderable medium as well as in a vacuum; the apparent influence of a medium on the observable force $\mathbf{F}_m$ is attributed to magnetization (magnetic polarization) of the medium and to magnetostriction. The mathematical theory is formally analogous to that of electrostatics, with the magnetizing force, or magnetic intensity, **H** replacing the electric intensity **E**.[17] It follows at once that the primary conclusion reached in Sec. 2.1 for electrostatic theory carries over to magnetostatics. *The equation under investigation, Eq. (2–15), is not generally valid and is anything but fundamental.* In fact, as will appear, the standing of Eq. (2–15) in magnetostatics is much less respectable than the standing of the corresponding Eq. (2–1) in electrostatics.

[15] In practice it will, of course, suffice to make the dimensions of the test-body small enough to eliminate multipole forces and to make the test-charge small enough to eliminate image-forces. These restrictions apply equally to the measurement of electric intensity in a vacuum or in a fluid.

[16] It may be observed that when Eq. (2–5) is known to be correct, the resultant force of electrical origin due to the action of $\mathbf{E}'$ on the polarization charges in, and immediately adjacent to, the body under consideration is canceled by the resultant mechanical force due to electrostriction in the fluid. See D. L. Webster, *Am. Physics Teacher* **2**, 149–151 (1934).

[17] It is, of course, possible to develop magnetic theory from the beginning in terms of electric currents—a procedure discussed in Sec. 2 and illustrated in detail in Outline II, Sec. 3.3. The two approaches lead to the same results.

Magnetostatics as traditionally developed in terms of the magnetic pole concept is a much clumsier subject than electrostatics. In the absence of a magnetic equivalent of conduction charge the roles of ρ_c and σ_c in electrostatics are taken over by the volume and surface densities of "permanent" magnetic pole strength, ρ_{m0} and σ_{m0}, respectively. These quantities are defined by the equations[17a]

$$\rho_{m0} = -\mathrm{div}\mathbf{M}_0, \quad \sigma_{m0} = \mathbf{n}\cdot\mathbf{M}_0, \qquad (2\text{–}17)$$

where $\mathbf{M}_0$ is that part of the magnetization $\mathbf{M}$ identified as "permanent." It follows that the important electrostatic case of conductors separated by a dielectric has no parallel in magnetostatics, and that the analog of Eq. (2–4) does not hold rigorously except in a physically unrealizable limit. If a magnet is immersed in a Class A magnetic fluid, that is, a homogeneous, isotropic, and magnetically linear fluid, the interior of the magnet is inaccessible to the fluid and is not field-free. Therefore the conditions for validity of the magnetic form of Eq. (2–4) are not satisfied. The case of small charged bodies also has no exact parallel; the closest approximation is realized by using slender rods of magnetically hard material subjected to suitable magnetic preparation. It is to such magnets, if to any, that the analogs of Eqs. (2–1), (2–4), and (2–5), *viz.*, Eq. (2–15), and

$$\left.\begin{aligned} &\text{(Gauss)} && \frac{dp'}{\kappa_m r^2}\mathbf{1}_r && (2\text{–}18a)\\ &\text{(Giorgi-Kennelly)} && \frac{dp'}{4\pi\mu_0\kappa_m r^2}\mathbf{1}_r && (2\text{–}18b)\\ &\text{(Giorgi-Sommerfeld)} && \frac{dp'}{4\pi\kappa_m r^2}\mathbf{1}_r && (2\text{–}18c)\end{aligned}\right\} = d\mathbf{H}'$$

$$\left.\begin{aligned} &\text{and (Gauss or Giorgi-Kennelly)} && \mathbf{H}'dp && (2\text{–}19ab)\\ &\text{(Giorgi-Sommerfeld)} && \mu_0\mathbf{H}'dp && (2\text{–}19c)\end{aligned}\right\} = d\mathbf{F}_m$$

should apply. Actually the validity of these equations in the idealized limiting case of permanent, infinitesimally thin magnets immersed in a large volume of Class A magnetic fluid can be verified without difficulty. They do not hold for magnets of other shapes, however, and Eq. (2–18) fails even for a needle-magnet if inhomogeneities in the medium have to be taken into account. The valid uses of Eq. (2–18) are so unimportant and the possibility of misuse is so great that its introduction into an elementary course is pretty clearly inadvisable.

Equation (2–19) in a form analogous to that of Eq. (2–14) may be used to introduce the magnetic intensity concept in an elementary course, but if it is so used the limitation to very thin slender magnets should be clearly pointed out.

Particular attention is called to the fact that the shape of a magnet has an important effect on the change in its field that is caused by immersing it in a magnetic fluid and also on the change in the force and torque experienced when it and other sources of magnetic field are immersed in a magnetic fluid.

A good introductory description of the effect of immersion on the field of a permanent magnet is provided by the elementary theory of the magnetic circuit. The field of a magnet may be regarded as the result of the action of a magnetomotive force through a reluctance. If the external reluctance is decreased by immersing the magnet in a paramagnetic fluid, the total flux increases, but the potential drop across the external part of the circuit decreases; that is, the external $\mathbf{B}$ increases but the external $\mathbf{H}$ decreases. In the limiting case of long thin needle-magnets magnetized longitudinally, the internal reluctance far exceeds the external and therefore determines the flux; consequently, the external flux is independent of the reluctivity of the fluid and the potential drop is proportional to the reluctivity, that is, the fluid leaves $\mathbf{B}$ unaffected and decreases $\mathbf{H}$ in the ratio $1/\kappa_m$. Conversely, in the limiting case of a thin disk-magnet magnetized transversely (magnetic shell) the internal reluctance is negligibly small and, therefore, the magnetomotive force acts directly on the external part of the circuit; consequently, the external potential drop is independent of the reluctance of the fluid and the flux is proportional to the permeability; that is, the fluid leaves $\mathbf{H}$ unaffected and increases $\mathbf{B}$ in the ratio κ_m.

This approach to the problem yields a generally correct account of the behavior of actual

[17a] Explicit definitions of magnetization $\mathbf{M}$ and magnetic induction $\mathbf{B}$ are given in Sec. 3.23.

magnets and indicates that the needle-magnet is an extreme case—interesting because it gives the closest magnetic analog of the point charge, *but no more representative of magnets in general than the opposite extreme, the disk-magnet.*

Elementary magnetic circuit theory also provides a rough description of the behavior when the change of external reluctance is produced by bringing ferromagnetic objects into the neighborhood—a case of much greater practical importance than that of immersion in a fluid. However, such a description must not be misinterpreted as a substitute for field theory. A magnet is not, in general, bounded by equipotential surfaces and lines of induction. Therefore, the internal and external reluctances are not quantities that can be defined precisely until the flux distribution has been found. Moreover, a change of external reluctance usually changes the internal flux distribution and therefore somewhat alters the internal reluctance. An ultimate analysis by field theory is therefore as necessary with this approach as with one based on poles.

The conclusions just indicated in terms of magnetic circuit theory can be restated in another form as follows: Let p_0 and p_t denote, respectively, the vacuum pole strength and the total, or net effective, pole strength of a magnet when immersed in a Class A paramagnetic fluid. (Here p_0 is assumed to be constant for a given magnet, whereas p_t, which includes the pole strength induced in the adjacent medium, depends on the medium. Appendix B contains a more detailed discussion of the ratio p_t/p_0 for real as well as ideal magnets.) Then our first conclusion is that in the case of an ideal needle-magnet $p_t = p_0/\kappa_m$. In the second place, since the external magnetic field of a magnetic shell is proportional to the product of the pole strength per unit area and the thickness, the induced pole density in the medium at the surface of an immersed magnetic shell is always negligible in comparison with the vacuum pole-density. Hence the pole strengths p_t and p_0 are equal for such a shell. The **H**-field of a magnetic shell is accordingly independent of the permeability of a surrounding fluid.[18]

The law of action and reaction is sufficient to establish the proposition that the differences between the effects of immersion on the fields of magnets of different shapes must be reflected in corresponding differences in the forces which the two classes of magnets experience when subjected to external fields and immersed.

To work out these differences we may take over from electrostatic theory[19] the force equation (2–11) whose magnetic equivalent is

$$\left.\begin{array}{l}\text{(Gauss or Giorgi-Kennelly)}\\ \qquad (\kappa_m)_A \displaystyle\int_X \mathbf{H}\,dp_t \qquad (2\text{–}20\text{ab})\\ \text{(Giorgi-Sommerfeld)}\\ \qquad \mu_0(\kappa_m)_A \displaystyle\int_X \mathbf{H}\,dp_t \qquad (2\text{–}20\text{c})\end{array}\right\} = \mathbf{F}_m.$$

Here, as in Eq. (2–11), the subscript A is used to denote a property of the Class A fluid; dp_t denotes an element of total pole strength due to both the permanent and variable components of the magnetization. The integration is to extend over the volume of the magnet X that experiences the force $\mathbf{F}_m$ and over the surface distribution of poles induced in the fluid adjacent to the surface of X.[20]

[18] Since the **H**-field of an electric current is unchanged by immersion in a magnetic fluid, the familiar equivalence between a linear current and a corresponding magnetic shell is undisturbed if shell and circuit are both "dunked" in magnetic fluid.

[19] A complete mathematical correspondence can be established between the computation of the resultant forces and torques on charged bodies in a Class A dielectric fluid as described in Appendix A and the computation of the corresponding forces and torques on magnets in a Class A magnetic fluid. In this correspondence **E** is replaced by **H**, the total electric charge density ρ_t is replaced by the total magnetic pole density $\rho_{mt} = -\text{div}\mathbf{M}$, and the conduction charge density ρ_c is replaced by the contribution of the "hard" polarization $\mathbf{M}_0$ to the total pole density. In Appendix A we replace the charged body X by an imaginary body X^* having the same total charge density but with a uniform relative dielectric constant equal to that of the surrounding Class A fluid. Similarly in the magnetic case we introduce an imaginary magnet having the same total pole density as X but provided with a "soft" polarization $\mathbf{M}_s = \mathbf{M} - \mathbf{M}_0$ which is proportional to **H**. By giving this substitute magnet a suitable susceptibility for soft polarization we can make its effective permeability the same as that of the surrounding Class A magnetic fluid. Thus the force equation (2–20) and a torque equation corresponding to Eq. (A–b) are readily established.

[20] There are a number of important equivalent expressions for the force and torque derivable from Eq. (2–20) and the magnetic analog of Eq. (A–b) of Appendix A by mathematical transformation. Assuming such transition

It is not permissible to include only one pole in the volume of integration, or in any way to fail to include the whole magnet and a thin layer of fluid around it. To see why, one may refer to Appendix A, where Eq. (2–11) is derived, and specifically to its use of the condition that the whole surface bounding the volume of integration is in a Class A fluid. As this condition is used there, the word *fluid* is as important as the uniform susceptibility implied in *Class A*.

On this account Eq. (2–20) and the corresponding torque equation do not establish the existence of forces concentrated at poles. Neither do the equations for force and torque in terms of Amperian currents (curl$\mathbf{M}$ of footnote 20) establish the existence of forces on the sides of the magnet, where the uncanceled Amperian currents are concentrated. On the contrary, we must assume that, aside from minor contributions due to the fluid, the bulk of the forces and torques acting on a magnet are spread throughout its volume along with the electrons whose spins account for the magnetization.

In problems having to do with the mechanical behavior of magnets either style of concentration may be assumed and either works as well as the other. But this is only because of the great rigidity of ferromagnetic materials; so it is appropriate that this rigidity, by contrast with the fluidity of the Class A fluid, is just what makes Eq. (2–20) inapplicable to anything less than a whole magnet.

In Eq. (2–20) we can substitute for the total intensity $\mathbf{H}$ the part $\mathbf{H}'$ due to magnetization or currents outside X. Finally, we express $\mathbf{H}'$ in terms of the magnetic induction $\mathbf{B}'$ which would have existed at the point under consideration if the magnet X were taken out of the fluid. In this way we obtain the equivalent equation

$$\left.\begin{array}{l}\text{(Gauss or Giorgi-Sommerfeld)}\\ \qquad\qquad \displaystyle\int_X \mathbf{B}'\,dp_t\end{array}\right\} = \mathbf{F}_m. \qquad (2\text{–}21ac)$$

$$\left.\text{(Giorgi-Kennelly)}\ \ \mu_0^{-1}\int_X \mathbf{B}'\,dp_t\right\} \qquad (2\text{–}21b)$$

Similarly the torque $\mathbf{L}_m$ acting on X can be expressed in the form

$$\left.\begin{array}{l}\text{(Gauss or Giorgi-Sommerfeld)}\\ \qquad\qquad \displaystyle\int_X \mathbf{r}\times\mathbf{B}'\,dp_t\end{array}\right\} = \mathbf{L}_m. \qquad (2\text{–}22ac)$$

$$\left.\text{(Giorgi-Kennelly)}\ \ \mu_0^{-1}\int_X \mathbf{r}\times\mathbf{B}'\,dp_t\right\} \qquad (2\text{–}22b)$$

These apparently simple equations are deceptive and must be used with caution. They are valid only in a Class A fluid—a restriction not immediately evident as in the parallel form, Eq. (2–20). Moreover, it is *not* permissible to replace $\mathbf{B}'$ in Eqs. (2–21) and (2–22) by the actual net magnetic induction $\mathbf{B}$. This is because within the magnet the product $(\kappa_m)_A\mathbf{H}$, which occurs in Eq. (2–20), is very different from $\mathbf{B}$, and there is no certainty that the difference would integrate to zero. Likewise, the $\mathbf{B}$ in the equations containing curl$\mathbf{M}$ in footnote 20 can be replaced by $\mathbf{B}'$, and then $\mathbf{B}'/(\kappa_m)_A$ by $\mathbf{H}'$; but it is *not* permissible for *this* $\mathbf{H}'$ to be replaced by $\mathbf{H}$.

Moreover, care must be used with the definitions of $\mathbf{H}'$ and $\mathbf{B}'$ as vectors which would have existed if the magnet were removed. Before this removal any magnetization induced by this magnet in other bodies would have to be *frozen into* them, so that it would contribute to $\mathbf{H}'$ and $\mathbf{B}'$.

Altogether, if Eqs. (2–21), (2–22) and their Amperian analogs are used, these hidden restrictions and those of Eq. (2–20) must all be kept in mind; but Eq. (2–20) and its Amperian analog are subject only to the more obvious restrictions, that the region of integration must include the whole magnet X and a surface film of a single surrounding Class A fluid. In any case one must

---

layers as are necessary to convert surface integrals into volume integrals and restricting ourselves to Gaussian units, we may write, for example:

$$\mathbf{F}_m = (\kappa_m)_A\int(-\mathrm{div}\mathbf{M})\mathbf{H}\,d\tau = [1/(\kappa_m)_A]\int[(\mathrm{curl}\mathbf{M})\times\mathbf{B}]\,d\tau,$$

$$\mathbf{L}_m = (\kappa_m)_A\int[\mathbf{r}\times(-\mathrm{div}\mathbf{M})\mathbf{H}]\,d\tau$$

$$= \frac{1}{(\kappa_m)_A}\int[\mathbf{r}\times\{(\mathrm{curl}\mathbf{M})\times\mathbf{B}\}]\,d\tau.$$

remember that the polarization of the surrounding medium alters the distribution of total pole strength and would, in fact, produce a net force and net torque on a completely nonmagnetic body immersed in a magnetic fluid and exposed to an external magnetic field.[21] For this reason one can draw simple conclusions from Eqs. (2–21) and (2–22) only in simple special cases.

Of prime interest is the case of an ideal needle-magnet subjected to an external field which is sensibly uniform in the small volume over which its effective pole strength is distributed. In this case it follows from the preceding equations that each pole can be considered subject to a concentrated force applied at the pole centroid whose value is expressible in terms of $\mathbf{B}'$ or $\mathbf{H}'$ as follows:

$$\left.\begin{array}{ll} \text{(Gauss)} & \\ \quad \mathbf{B}'p_t=\mathbf{B}'p_0/(\kappa_m)_A=\mathbf{H}'p_0 & (2\text{–}23a) \\ \text{(Giorgi-Kennelly)} & \\ \quad \mu_0{}^{-1}\mathbf{B}'p_t=\mathbf{B}'p_0/\mu_0(\kappa_m)_A=\mathbf{H}'p_0 & (2\text{–}23b) \\ \text{(Giorgi-Sommerfeld)} & \\ \quad \mathbf{B}'p_t=\mathbf{B}'p_0/(\kappa_m)_A=\mu_0\mathbf{H}'p_0 & (2\text{–}23c) \end{array}\right\}=\mathbf{F}_m.$$

Thus Eq. (2–19) holds as a satisfactory approximation for slender needle-magnets provided we identify dp with an element of the fixed vacuum pole-strength p_0. There is, consequently, no serious misrepresentation of the facts if we choose to identify the theoretically defined magnetic intensity $\mathbf{H}'$ in a Class A fluid (see Sec. 2.5) with the force-per-unit-pole-strength on a concentrated pole at the point in question.

Some authors[22] prefer to stress total pole-strength p_t rather than vacuum pole-strength p_0. If this is done the force equation is written as $\mathbf{F}_m=p_t\mathbf{B}'$ and the factor κ_m in the Coulomb's law equation (2–15) is thrown into the numerator. This practice may have advantages, but it introduces a variable where a corresponding invariant quantity could be used and forces a corresponding alteration in electrostatic theory if the usual analogy between electrostatics and magnetostatics is to be maintained. It is misleading indeed to indicate that $\mathbf{F}_m=p_t\mathbf{B}'$ is the *only* correct form.

[21] Moreover, a large external field, like that obtained when opposite poles of two different magnets are placed in contact, may be expected to bring about hysteresis effects not considered in elementary theory.

[22] For example, John A. Eldridge, *Am. J. Physics* **15**, 390 (1947), **16**, 327 (1948); E. Hallén, *Trans. Roy. Inst. Tech., Sweden*, No. 6 (1947).

In conclusion it should be pointed out that, if magnetostatics is developed from the law of interaction between current elements, there is the same objection to the premature introduction of κ_m as when we start from Coulomb's law for magnetic poles. The equation

$$\left.\begin{array}{ll} \text{(Gauss)} & \dfrac{I}{c}d\mathbf{s}\times\oint\left[\dfrac{I'}{c}d\mathbf{s}'\times\dfrac{\mathbf{1}_r}{r^2}\right] \quad (2\text{–}24a) \\ \text{(Giorgi)} & \dfrac{\mu_0}{4\pi}Id\mathbf{s}\times\oint\left[I'd\mathbf{s}'\times\dfrac{\mathbf{1}_r}{r^2}\right] \quad (2\text{–}24b) \end{array}\right\}=d\mathbf{F}_m$$

is analogous to Eq. (2–16) for this approach. It gives the force exerted on a current element $Id\mathbf{s}$ by a complete circuit composed of elements $I'd\mathbf{s}'$. The same equation with a factor κ_m inserted in the left-hand member is supposed to take into account immersion in a Class A magnetic fluid and is the analog of Eq. (2–15). This modified equation is not fundamental and holds only under restrictions that should be clearly specified if it is to be used at all.

The basically correct way to develop the theory is to begin with Eq. (2–24). The effect of a magnetic medium is then explained as due to the magnetization of the medium and the associated distribution of uncanceled Amperian currents, whose density is

$$\left.\begin{array}{lll} \text{(Gauss)} & c\ \text{curl}\mathbf{M} & (2\text{–}25a) \\ \text{(Giorgi-Kennelly)} & (1/\mu_0)\ \text{curl}\mathbf{M} & (2\text{–}25b) \\ \text{(Giorgi-Sommerfeld)} & \text{curl}\mathbf{M} & (2\text{–}25c) \end{array}\right\}=\mathbf{J}_m.$$

The **B**-field of the magnetization is then the field of the current distribution $\mathbf{J}_m$, while the **H**-field is computed from the corresponding distribution of pole strength.

When a Class A magnetic medium fills all regions where there is a magnetic field, the magnetization produces no poles and the **H**-field is the same as without the medium; **B** is therefore multiplied by the factor κ_m due to the medium. Under these conditions the vector potential differs from that without the medium only by a factor κ_m and the apparent forces between circuits are correctly given by Eq. (2–24) with κ_m inserted in the numerator. The situation in which a single Class A magnetic medium fills all space where there is a field, is not of great importance in practice, however.

In elementary courses Eq. (2–24) is not likely to be used; but the simpler equation for the force of attraction between parallel wires is practically always discussed. The introduction of a factor κ_m in this equation involves much the same dangers as its introduction in the more general equation.

To summarize: The magnetic equation (2–15) does not constitute a satisfactory basis for the theory of magnetostatics. Like its electrostatic analog, Eq. (2–1), it has only a limited validity and its applications are not of great practical importance. Since their use has led, and is likely to lead, to misconceptions on the part of the students, the committee recommends: *that Eqs. (2–1) and (2–15) be omitted altogether in elementary courses;*[23] *that Coulomb's and Ampere's laws be used only in the basic forms of Eqs. (2–2), (2–16), and (2–24); and that the effect of an intervening medium be treated in such a manner as to insure an understanding of the physical processes occurring and an appreciation of the fact that the problem is, in general, not simple.*

2.3 Polarization as a Basic Concept

A second important issue in connection with the teaching of electrostatics relates to the relative emphasis to be placed on the polarization and displacement vectors. The displacement vector **D** has traditionally had the major emphasis, and this vector is convenient for the formulation of the general partial differential equations of the electromagnetic field in linear media. On the other hand, from our present-day point of view, the polarization vector **P** is simpler and more fundamental as a physical concept. We are bound to explain the behavior of dielectrics in terms of the electrical structure of matter by means of the concept of polarization. In Maxwell's day **D** was fundamental because his interpretation of dielectric behavior was altogether different from ours, but from the physical point of view it must now be regarded as of secondary importance.

The treatment of dielectrics in introductory textbooks of general physics usually does not extend much beyond the application to parallel-plate condensers. It is easy to give a physical explanation of dielectric action for this special case in terms of the concept of polarization. The displacement vector is not required for this purpose and need not be introduced at all in such a text unless the author wishes to use it as a preparation for the magnetic induction **B**.

In more advanced textbooks the usefulness of the vector **D** is unquestionable, but in the opinion of the majority of the committee there is a great advantage even in such books in stressing the primary physical importance of **P**.

2.4 Definitions of the Electric Vectors at Points Inside Ponderable Media

The problem of defining **E**, **D**, **B**, and **H** at points inside ponderable media has long been a thorn in the side of the teacher of elementary physics and to a large extent the issue has been evaded in ways that are unsatisfactory.

Some of the definitions do not define; others are given as if they were general when actually they are applicable only to special cases. We see no objection in the construction of an elementary textbook to the use of definitions of limited scope in the place of general definitions, but in such cases the reader should be very clearly warned of the limitation.

We believe that confusion frequently arises in the minds of teachers and students alike as a result of the common failure to distinguish between two different ways in which we use the basic terms of our science. For example, we use the term "electric intensity" and the symbol **E** to denote what we conceive to be an objective physical reality susceptible of measurement in fact, or in principle, by the force-per-unit-test-charge technique. We use the same phrase and symbol to denote the corresponding theoretical concept, namely, the mathematical entity **E** as it appears in theoretical computations such as the evaluation of the capacitance of a condenser.

These two functions of the term are, of course, intimately related. The rules by which we use the theoretical concept in our calculations are formulated on the basis of experimental measurements in such fashion as to make the drawing of a distinction between the two meanings appear at first glance to be artificial and unnecessary. Nevertheless, the committee believes that a frank

[23] In making this recommendation the Committee follows the lead of Professor Murdock, whose paper (Ref. 1) led to the present inquiry.

recognition of a distinction between what we may describe as the experimental meaning of a physical quantity and the meaning of the corresponding theoretical concept can be very helpful.[24] One meaning is defined by a set of laboratory operations, while the other is defined by the mathematical rules governing the paper-and-pencil operations of problem-solving. The distinction is similar to that between physical dots and lines on paper and the ideal points and lines of the geometer.

In the past physicists have tended to ignore this difference in meaning because they have adopted a naive common-sense philosophy which identified the quantity under consideration in each case with a single objective property of the external world. Such a view does not adequately recognize the essentially tentative character of physical theories. Moreover, as we endeavor to show in the following paragraphs, at certain points this common-sense point of view muddies the business of formulating theories.

Consider the electric intensity vector. Direct measurements of electric intensity as force-per-unit-test-charge are infrequent and crude. On the other hand, the theoretical tool **E** is sharply defined as one element of a conceptual scheme containing many others, such as the polarization vector, electric charges, space coordinates, conducting surfaces, etc. It is a vector point-function used with other functions in the scheme to give solutions of Maxwell's equations. The elements of uncertainty which characterize experimental measurements have no place in the theory because the theory is a mental construction.

We are inclined to postulate a structural relation between elements of physical reality corresponding exactly to the structural relations of a useful theory, like the classical theory of electromagnetism. The need for a quantum theory of electrodynamics clearly indicates, however, that the postulated exact correspondence does not exist, at least for the classical theory of electricity and magnetism.

There is, of course, a correspondence between the elements of the conceptual scheme used in any domain of physics and certain measurable physical quantities that must be observed in setting up a specific theoretical problem from data for a specific experimental situation, as well as in comparing the theoretical result with the outcome of the experiment. Rarely, however, are all the elements of the theoretical scheme susceptible of direct comparison with their objective physical correlates. We may and do use theoretical concepts which have no direct physical correlates at all, such as the wave functions of quantum mechanics.

The distinction we are emphasizing becomes important for the beginner when the discussion relates to the interior of a solid medium, where the electromagnetic vectors are never measured.[25] It is true that the Kelvin-cavity definitions provide us with mental experiments which give the semblance of physical reality to the theoretical concepts **E**, **D**, **H**, and **B**. These definitions are never converted into concrete physical experiments, however, and would involve insuperable difficulties in asymmetric crystals, in ferromagnetic materials exhibiting hysteresis, and in cases in which the vectors change rapidly in time, as in the passage of light waves through matter. From the point of view of the Lorentz electron theory, the electric and magnetic vectors **E** and **B** in material media are space averages of violently fluctuating "microscopic" field vectors and atomic magnitudes that are never susceptible of direct measurement. Hence, *the macroscopic vectors* **E**, **D**, **H**, **B**, *and the electric and magnetic polarizations should be defined for points in solid bodies as theoretical tools and not as directly measurable physical quantities. The definitions should serve to show the student how the various vectors are to be used in calculations and it should be made clear to the student that this is the purpose of the definitions.*

With these considerations in mind the committee recommends for the definition of the electric vectors a procedure somewhat as follows:

First discuss the electrostatics of free space, defining **E**

[24] The distinction has a close similarity to that drawn by F. S. C. Northrop between "concepts by intuition" and "concepts by postulation," although the experimental meaning of such a concept as the electric intensity does not make it a concept by intuition. See Northrop's *Logic of the sciences and the humanities* (The Macmillan Company, New York, 1947), esp. pp. 60, 82, 83.

[25] Of course, we *do* make measurements of the average values of these vectors; for example, the measurement of the magnetic flux through a solenoid with an iron core determines the average value of B in the core.

by the usual force-per-unit-test-charge method. Introduce dielectrics by an account of the experimental behavior of condensers whose plates are separated by such media. Mention some other experiments of interest in connection with dielectrics, such as the observed change in an electric field due to the insertion of a block of dielectric into the field, the measurement of the force of attraction of a point charge for an uncharged block of dielectric, or the force between two charged bodies immersed in a fluid dielectric. Then confront the class with the problem of developing a theory which, if possible, shall give a quantitative prediction of many experimental results and which shall relate the observations in a reasonable way to other information regarding the probable nature of matter. Such a theory can reasonably consist of a *model* of the dielectric based on familiar concepts and forming a basis for the computation of quantities which are to be compared with experimental results.

As a first step toward the construction of a theory, introduce the idea that matter contains distributions of positive and negative charge which are very dense, but which ordinarily cancel one another, and which are bound together in most cases by strong "elastic" (linear) forces. Note that from the atomic point of view these distributions are undoubtedly discontinuous, but for the purpose in hand can be treated as continuous.[26] Under the influence of an electric field a small motion of positive charge density, ρ^+, relative to negative charge density, ρ^-, may be expected. Such a motion would create a net surface-density of bound charge (polarization charge) at the surface of the medium and under certain circumstances could create a non-vanishing net volume-density of polarization charge. *Postulate* that the electric intensity **E** outside the dielectric, where it is directly measurable, is to be computed from the total charge, including both free and bound parts, by the methods used in the absence of dielectrics. Note that although we can not directly explore the electric field inside a solid dielectric, we can still *define* the electric intensity **E** for points where it can not be measured as a vector computed from the charge distribution by the same rules that apply outside the dielectric. The emphasis to be laid on the distinction between the two meanings of the term "electric intensity" will depend on the level of the course and the bias of the teacher.

Next *define* the polarization vector **P** in terms of the model of the dielectric as electric moment per unit volume due to the relative displacement of positive and negative bound charges. Its value at any point A is equal to the product of the positive charge density and the vector displacement of positive charge relative to negative charge, both evaluated at A. From this definition show that the polarization of the medium must cause the flow of bound charge across every surface element dS in the direction of the positive normal in amount $P_n dS$, where P_n is the component of **P** in the direction of the positive normal. It follows that at the exterior surface of a dielectric there should be a net bound surface-charge-density equal to the component of **P** at the surface in the direction of the exterior normal. If the mathematical level of the course permits, show that the negative divergence of **P** (if any) is the net volume-density of bound charge. Thus the vector **P** determines the complete distribution of net bound charge, or polarization charge.

At this point the conceptual scheme devised for the interpretation of the behavior of dielectrics consists of four essential elements, *viz.*, the density of conduction charge ρ_c, the polarization vector **P**, the net density of the polarization charge, and the electric intensity **E**. The first two quantities determine the second two, but none of the quantities is directly measurable in the interior of the dielectric. The four concepts *belong to the model.*

In applying the theory to special cases an additional assumption is needed in order to create a mathematical problem which has a definite solution. Consider the case of a parallel-plate condenser large enough so that edge effects can be neglected. Let the space between the plates be filled with an ideal homogeneous isotropic dielectric. In this case the density of conduction charge can be assumed equal to zero except on the surfaces of the plates, where it must have a uniform value σ_c (positive plate), or $-\sigma_c$ (negative plate). In conformity with the experimental proportionality of charge and potential difference the polarization **P** of the model is assumed to be proportional to **E** and in the same direction. This assumption completes the essential features of the model. Show that under these circumstances the model predicts that the dielectric will increase the capacitance of the condenser by the factor $1+4\pi\chi_e$ (Gauss) or $1+\chi_e$ (Giorgi), where the electric susceptibility χ_e is defined as the dimensionless constant of proportionality in the relation

(Gauss) $\mathbf{P}=\chi_e\mathbf{E}$ (2–26a)

or (Giorgi) $\mathbf{P}=\epsilon_0\chi_e\mathbf{E}$ (2–26b)

Define the relative dielectric constant κ_e by the equation

(Gauss) $\kappa_e=1+4\pi\chi_e$, (2–27a)

or (Giorgi) $\kappa_e=1+\chi_e$. (2–27b)

Then take one of the other special problems of the elementary theory of dielectrics and show that if the dielectric is homogeneous, isotropic, and linear, the model and the information contained in the specification of the experimental conditions suffice to yield a definite scheme for computing the result of the observations.

Attention should be called to the fact that the assumption that **P** is proportional to **E** does not give a satisfactory model for many dielectrics.

In conclusion, note the fact that a small charged body immersed in a fluid dielectric should experience three kinds of forces, *viz.*, electric forces due to its direct interaction with other conduction charges in the neighborhood, electric forces due to interaction with the electric dipoles which the dielectric contains, and a mechanical force due to the asymmetric pressure built up in the fluid by electrostrictive action. Note that the resultant of these forces is $q_c\mathbf{E}$ so that in

[26] At the pleasure of the instructor the charges may be treated as discontinuous during the early stages of the discussion, the transition to continuous charge densities being deferred as long as possible.

principle **E** can be measured in a fluid dielectric by the same technique as in a vacuum.[27]

At this point the vector **D** can be introduced and defined in terms of **E** and **P** by the equation

$$\text{(Gauss)} \qquad \mathbf{D}=\mathbf{E}+4\pi\mathbf{P}, \qquad (2\text{–}28a)$$

$$\text{or (Giorgi)} \qquad \mathbf{D}=\epsilon_0\mathbf{E}+\mathbf{P}. \qquad (2\text{–}28b)$$

In a Class A dielectric in which **P** is proportional to **E** this definition leads at once to the simple relation $\mathbf{D}=\kappa_e\mathbf{P}$, or $\mathbf{D}=\epsilon_0\kappa_e\mathbf{P}$, according to the unit scheme adopted.

Call attention to the fact that unlike **E** the vector **D** is not derivable from a potential except in special cases where it has no component parallel to any dielectric interface in the field. When the mathematical properties of **D** have been sufficiently developed, the Kelvin-cavity interpretations of **E** and **D** may be given to show how far one can go in the way of giving a sense of physical reality to the vector concepts of the mathematical theory.

Some authors have preferred to introduce **D** at an early stage as a vector whose divergence is proportional to the conduction charge density and which is related to **E** by the rule, $\mathbf{D}=\epsilon\mathbf{E}$. Since curl**E** vanishes identically, curl**D** reduces to the vector product $\nabla\epsilon\times\mathbf{E}$. Thus curl**D** vanishes in a homogeneous isotropic medium, and in a transition layer between one dielectric and another if **E** happens to be normal to the interface. It also vanishes automatically at the interface of a dielectric and a conductor in which there are no currents. There are, accordingly, a number of problems in which **D** is irrotational and therefore correctly given by the simple formula

$$\text{(Gauss)} \qquad \mathbf{D}=\Sigma q\mathbf{1}_r/r^2, \qquad (2\text{–}29)$$

where $\mathbf{1}_r$ denotes a unit vector along the radius from the source-point to the field-point. Although this procedure is open to no logical objection if the severe restrictions on the formula are made clear, there is great danger that the use of Eq. (2–29) in the early stages of an elementary or intermediate course will lead to a serious misconception in the minds of many students. Hence *the committee regards such use of Eq. (2–29) as undesirable.*

2.5 Concerning the Definitions of the Magnetic Vectors and Some Related Matters

The business of defining the magnetic vectors **B**, **M**, and **H** in solid magnetic media involves the same need for distinguishing between theoretical and experimental meanings that we have discussed in connection with the electric vectors **E**, **P**, and **D**. The definitions of **B**, **M**, and **H** in terms of laboratory operations have significance only for media in which the operations can actually be performed. The theoretical meanings, valid for all media, are based on mathematical relations, or paper-and-pencil operations, that show how these vectors are to be used in calculations based on the classical theory. In this respect, however, the magnetic vectors present no new pedagogical problems.[27a]

A different problem connected with the definition of **M** has arisen, however, as the result of the possibility of interpreting the elementary magnets of the "molecular" theory either as pairs of equal and opposite poles, or as Amperian current loops. The ambiguity is purely formal, since nearly everyone agrees that the magnetic poles are mathematical fictions and that the Amperian currents correspond closely to present conceptions of the physical "reality."[28] Each interpretation is allied to a method of calculation and the two methods are mathematically equivalent. Nevertheless, considerable divergences in teaching practice and a recently developed disagreement as to the best choice of unit and dimensions for magnetic polarization in the Giorgi scheme are consequences of this double interpretation.

There are, in fact, three different, closely related issues which should be considered in this connection, namely, (a) the relative emphasis to be placed on magnetic poles and Amperian currents as initial concepts in the theory of magnetism, (b) the relative emphasis to be placed on the correlation of **H** with **E** (parallel correlation) and on the correlation of **B** with **E** (antiparallel correlation), and (c) the definition of **M** (choice of dimensions and unit in the Giorgi system). As will appear in Secs. 2.6 and 2.7, a decision to

[27] See Sec. 2.1, Eqs. (2–5) and (2–14).

[27a] The explicit working out of these definitions with their appropriate background is carried through in Sec. 3.25 of Part III.

[28] Yet too much emphasis should not be placed on the supposed reality of Amperian currents. Ferromagnetic polarization is primarily due to electron-spin magnetic moments. Few of those familiar with the rules of behavior of electron spin would consider them a very satisfactory justification for calling Amperian currents "real." Still more serious doubts regarding a naive acceptance of the reality of Amperian currents might be raised by a consideration of neutron magnetic moments.

introduce the subject of magnetism by way of the magnetic-pole concept leads almost inevitably to emphasis on **H** and its correlation with **E** and, if Giorgi units are to be used, to preference for the Kennelly definition of **M**. On the contrary, a decision to introduce magnetism by the Amperian current concept leads almost inevitably to emphasis on **B**, to the correlation of **B** with **E**, and to preference for the Sommerfeld definition of **M**. Thus, the three issues reduce to a single choice between what we may refer to as the traditional, or magnetic pole, approach to the study of magnetism and the Amperian current approach.

In discussing this broad problem the Coulomb's Law Committee has found itself confronted with matters of taste and personal judgment as well as with matters of logical rigor. Although we have no mandate to render opinions on such issues, the confusion created by the two definitions and the highly partisan arguments on one side and the other call for clarification. Under these circumstances we here undertake to present the chief arguments on both sides, not so much to assist readers to the best possible final choice of approach as to give them, if possible, a fuller realization of the implications of the choice and a clearer conception of the possibilities of improved instruction by either technique.

2.6 Amperian Currents versus Magnetic Poles

The traditional approach to magnetism (see Outline I of Part 3) is by way of point poles, distributed poles, and distributed dipoles. After the magnetic fields of permanent magnets have been described on the basis of these concepts it is explained that current loops produce magnetic fields that are indistinguishable at large distances from the fields of small magnets. Amperian currents are then introduced as the equivalent of static dipoles. The magnetic fields of conduction-current circuits are introduced by means of the somewhat artificial concept of the magnetic shell.[29] By this device one can ultimately derive Ampere's law [Eq. (3–12)] and make both scalar and vector potentials available for the computation of magnetic fields.

This familiar procedure starts magnetic theory along lines to which the student has been introduced by the previous study of electrostatics and the polarization of dielectrics. It makes a maximum use of familiar mathematical tools and postpones vector products as long as possible.

The alternative Amperian approach (see Outline II of Part 3) is not so well standardized. It deals with the magnetic fields of currents in free space without the assistance of magnetic poles or permanent magnets. It starts from the observed forces between parallel currents rather than from Coulomb's law for magnetic poles, and defines the magnetic induction **B** in free space in terms of the torque on an exploring coil, or in terms of the force exerted on a straight current-carrying wire. Emphasis is laid from the beginning on important contrasts between magnetic and electric fields such as the fact that like currents attract one another although like charges repel.

When magnetic materials are taken up, the process of polarization by an external field is interpreted from the beginning as the rotation of Amperian-current loops due to the attraction of parallel currents and the repulsion of those that are antiparallel. The field of a magnetized bar of iron is referred, not to poles which accompany magnetization, but to net uncanceled Amperian currents on its lateral surface. In this way the discussion of the fields of currents in a vacuum is unified with the treatment of magnetic materials.

The arguments in favor of these two different strategies are various. Our discussion takes up first the simpler and more obvious considerations, then those that are not so obvious. Since "the proof of the pudding is in the eating" the final section of the discussion is contained in the two outlines for teaching given in Part 3. The remainder of this section expounds more fully the contrasts between the two approaches. The next section, 2.7, takes up the argument from the practical point of view of the classroom.

The traditional method of introducing the subject of magnetism describes the forces on poles in terms of **H** [see Eq. (2–19)]; the Amperian method describes forces on currents in

[29] Since transversely magnetized sheets of metal can now be manufactured, these shells are not so artificial as they once seemed to be.

terms of **B**. Hence the method of calculation used in the traditional approach makes **H** the analog of the electrostatic force **E**, but the Amperian method makes **B** the analog of **E**. These correlations are initially dictated by the fact that in a Class A magnetic medium the observable force on a concentrated magnetic pole is most conveniently expressed as $\mathbf{H}'p_0$, where $\mathbf{H}'$ is the external field and p_0 is the vacuum pole-strength [see Eq. (2–23)], whereas the force on a current-carrying wire of length L is $L\mathbf{I}\times\mathbf{B}'$ [see Eq. (3–11), where the prime is omitted as unnecessary in view of the context]. The correlations are reinforced by the satisfactory way in which they can be developed. Therefore, an essential part of the problem of choosing between the two pedagogical approaches is to see how the corresponding correlations work out.

The two schemes have many corresponding features. In the older one emphasis is laid on the identical form of the Coulomb law for charges and poles in a vacuum. In the newer one, as indicated above, a similar emphasis is laid on the contrast between the repulsion of like charges and the attraction of like currents. This contrast is conveniently formulated in terms of the equations for the forces on parallel wires of length L carrying, in the one case, charges Q and Q' per unit length and, in the other, currents I and I'. Thus

Electrostatic

(Gauss) $F_e=+(2QQ'L/r)$; (2–33a)

or (Giorgi) $F_e=+(QQ'L/2\pi\epsilon_0 r)$; (2–33b)

Magnetic

(Gauss) $F_m=-(2II'L/c^2r)$; (2–34a)

or (Giorgi) $F_m=-(II'L/2\pi\eta_0 r)$. (2–34b)

Here η_0 is a symbol introduced for the reciprocal of the usual μ_0 in order to remove an apparent contrast that is not of an essential character. The + and − signs, however, represent a fundamental contrast which we designate as *essential antiparallelism*, because the signs cannot be made alike without spoiling the symmetry of the definitions of F_e, F_m, I, I', or Q, Q'. Antiparallelism of this sort is characteristic of the Amperian correlation just as similarity, or parallelism, is characteristic of the traditional correlation.

Consider the lines of force. In the traditional approach lines of **H** are introduced by means of permanent magnets and in the absence of currents shown to have properties similar to those of the lines of **E**. In the Amperian approach, using electric currents and lines of **B**, similarity is largely replaced by contrast.[30] Thus electrostatic lines terminate on charges, but the lines of **B** form closed loops linked with currents (conduction, or Amperian, or both). In the teacher's mind, assuring him of the permanent value of these contrasts are the advanced equations:

Electrostatic

$$\mathrm{curl}\mathbf{E}=0, \qquad (2\text{–}35)$$

(Gauss) $$\mathrm{div}\mathbf{E}=4\pi(\rho_c+\rho_p), \qquad (2\text{–}36a)$$

or (Giorgi) $$\epsilon_0\,\mathrm{div}\mathbf{E}=\rho_c+\rho_p, \qquad (2\text{–}36b)$$

Amperian Magnetostatic

$$\mathrm{div}\mathbf{B}=0, \qquad (2\text{–}37)$$

(Gauss) $$\mathrm{curl}\mathbf{B}=4\pi(\mathbf{J}_c+\mathbf{J}_m), \qquad (2\text{–}38a)$$

or (Giorgi) $$\eta_0\,\mathrm{curl}\mathbf{B}=\mathbf{J}_c+\mathbf{J}_m. \qquad (2\text{–}38b)$$

Here, evidently, the antiparallelism takes the form of an exchange of the curl and divergence operators.

The corresponding equations exhibiting the parallelism of the traditional approach are:

Traditional Magnetostatic

$$\mathrm{curl}\mathbf{H}=0 \qquad (2\text{–}39)$$

(Gauss) $$\mathrm{div}\mathbf{H}=4\pi(\rho_{m0}+\rho_{ms}), \qquad (2\text{–}40a)$$

(Giorgi-Kennelly)

$$\mu_0\,\mathrm{div}\mathbf{H}=\rho_{m0}+\rho_{ms}, \qquad (2\text{–}40b)$$

(Giorgi-Sommerfeld)

$$\mathrm{div}\mathbf{H}=\rho_{m0}+\rho_{ms}. \qquad (2\text{–}40c)$$

Here ρ_{m0} and ρ_{ms} denote, respectively, the volume density of magnetic pole-strength due to perma-

[30] There are many ways of locating and defining the lines of **B** in lectures. Some of them will be discussed in more detail in Part 3, Outline II. Here we may note one definition chosen for contrast with almost any definition of the lines of **E**, namely: "A magnetic line of force in free space is a line along which a wire may lie and feel no force, regardless of the current it carries." The paradoxical aspect of this definition makes it take hold of the student's mind and helps to keep the class alert.

nent polarization and to "soft" polarization (see definition of ρ_{mt} in footnote 19).

The teacher must, of course, bear in mind the limitations of these equations and of the analogies which they describe. The curl equations are limited to cases in which the fields are not time-dependent. Equation (2–39) is restricted to cases in which there are no conduction currents and the division of ρ_m into ρ_{m0} and ρ_{ms} in Eq. (2–40) to cases in which the relation between **M** and **H** conforms to the linear equation

(Gauss or Giorgi-Sommerfeld)

$$\mathbf{M}=\mathbf{M}_0+\chi_m'(\mathbf{H}-\mathbf{H}_0) \qquad (2\text{–}41a, c)$$

(Giorgi-Kennelly)

$$\mathbf{M}=\mathbf{M}_0+\mu_0\chi_m'(\mathbf{H}-\mathbf{H}_0). \qquad (2\text{–}41b)$$

(See footnote 19 and Eq. (B-8) of Appendix B). Although the range of the parallel correlation of the traditional approach can be extended by the introduction of magnetic shells as the equivalent of linear currents, it will be seen that the parallel correlation is basically much more restricted in scope than the antiparallel one.

In describing polarizable media the traditional approach continues with parallelism, the Amperian with antiparallelism. Both explain the electrostatic effect of a glass plate inserted in an air condenser as a partial counterbalancing of the effects of conduction charges by unlike polarization charges attracted toward them. The traditional approach can use a parallel interpretation of the effect of a soft iron sheet placed between the poles of a permanent horseshoe magnet in reducing its external field. The Amperian approach, on the other hand, can introduce soft iron or other ferromagnetic material as a plunger in a solenoid, showing how it reinforces the external field and explaining the effect as the result of the attraction of Amperian currents of like sign by the conduction currents which surround the plunger.

On the basis of these explanations both approaches may define **D** by the equation

(Gauss) $$\mathbf{D}=\mathbf{E}+4\pi\mathbf{P}, \qquad (2\text{–}42a)$$

but the traditional approach now correlates **H** with **E**, whereas the Amperian correlates **B** with **E**. Thus

Traditional

(Gauss) $$\mathbf{B}=\mathbf{H}+4\pi\mathbf{M}, \qquad (2\text{–}43a)$$

Amperian

(Gauss) $$\mathbf{H}=\mathbf{B}-4\pi\mathbf{M}, \qquad (2\text{–}44a)$$

where the sign reversal of **M** in Eq. (2–44a) is the characteristic feature of antiparallelism.

To write either of these magnetic equations in Giorgi units, one must choose between the two conflicting definitions of **M** mentioned in footnote 7 and in Sec. 2.5. One of these, proposed by Kennelly and adopted by Harnwell and others,[31] corresponds to a definition of the magnetic moment of a plane current loop of area S as $\mu_0 IS$; the other, proposed by Sommerfeld and adopted by Stratton and others,[32] corresponds to the definition IS. The traditional equation (2–43) takes its simplest form in Giorgi-Kennelly units; the Amperian equation (2–44) in Giorgi-Sommerfeld units.[33] We accordingly write the Giorgi forms as follows:

(Giorgi) $$\mathbf{D}=\epsilon_0\mathbf{E}+\mathbf{P}, \qquad (2\text{–}42b)$$

Traditional

(Giorgi-Kennelly) $$\mathbf{B}=\mu_0\mathbf{H}+\mathbf{M}, \qquad (2\text{–}43b)$$

Amperian

(Giorgi-Sommerfeld) $$\mathbf{H}=\eta_0\mathbf{B}-\mathbf{M}. \qquad (2\text{–}44c)$$

The introduction of η_0 for $1/\mu_0$ is needed here again to give the antiparallel equations maximum symmetry.

For mechanical forces in polarizable media the traditional approach compares poles with charges as usual, while the Amperian approach contrasts

[31] A. E. Kennelly, Comité International des Poids et Mesures, *Procès-Verbaux des Séances* **17**, 1935; G. P. Harnwell, *loc. cit.*

[32] A. Sommerfeld, "Dimensionen der elektromagnetischen Grössen," *Zeits. f. Tech. Physik*, **16**, 420 (1935); J. A. Stratton, *loc. cit.*

[33] In *both* schemes the relations between magnetic moment, magnetization, magnetic pole-strength, and magnetic pole-density are the same. To convert any of these quantities from Sommerfeld units to Kennelly units we can start with pole-strength. For this the Gaussian unit is the (abampere×cm), the Sommerfeld unit is the (ampere ×meter), and the Kennelly unit is the (volt×second). The ratio of the Kennelly unit to the Sommerfeld unit is μ_0, or $4\pi\times10^{-7}$ henry/meter. The defining equation for susceptibility in the Sommerfeld scheme is $\mathbf{M}=\chi_m\mathbf{H}$, but in the Kennelly scheme it is $\mathbf{M}=\mu_0\chi_m\mathbf{H}$, which means that the values of χ_m and κ_m are the same for both.

currents with charges; so we can expect parallelism of signs in the one case and antiparallelism in the other. Equation (2–10) of Sec. 2.2 gives the force density resulting from electrostatic polarization as (Gauss) $-(E^2/8\pi)\,\mathrm{grad}\kappa_e$, or (Giorgi) $-\frac{1}{2}\epsilon_0 E^2\,\mathrm{grad}\kappa_e$. Corresponding to this, the traditional approach writes $-(H^2/8\pi)\,\mathrm{grad}\kappa_m$, or $-\frac{1}{2}\mu_0 H^2\,\mathrm{grad}\kappa_m$. In the Amperian approach, using B, η_0, and a variable relative reluctivity $\gamma_m = 1/\kappa_m$, the same term takes the form $+(B^2/8\pi)\,\mathrm{grad}\gamma_m$ or $+\frac{1}{2}\eta_0 B^2\,\mathrm{grad}\gamma_m$. That is, each approach leads to the sign which one would expect on the basis of its physical interpretation of magnetization.

Questions are often asked about what the field "really is" within a polarizable medium. As noted in Sec. 2.4 such questions have a preliminary answer in the conclusions of the Lorentz electron theory which make **E** the average of the "microscopic" electric field and **B** the average of the corresponding magnetic field. The Lorentz interpretations must be taken with a grain of salt since they antedate electron spin and the Heisenberg uncertainty principle. They harmonize very well with the Kelvin-cavity definitions, however, and the latter seem to have a certain permanent significance. So it may be well to point out that in the mental operation of cutting a drill-hole for the measurement of **E** one introduces no appreciable polarization charges. Just so, the cutting of a slot for the measurement of **B** introduces no appreciable uncanceled Amperian currents. So it is reasonable to identify these vectors with the average values of microscopic fields before making the cuts. It may be observed at the same time that if at this point one wishes to stick to poles, rather than currents, as the cause of magnetization, one should properly identify the **H** measured in a longitudinal drill-hole with the average initial microscopic field.

Finally, it may be well to note briefly what becomes of the correlations when we go beyond the limitations noted in connection with Eqs. (2–35) to (2–40). Going far beyond, into relativity, the Lorentz transformations deal with **E** and **B** in combination as one entity, and with **D** and **H** as another. Moreover, the four-dimensional vector analysis appropriate to relativity gives us Maxwell's equations for curl**E** and div**B** as parts of one vector equation, and those for curl**H** and div**D** as parts of another. In short, relativity confirms the correlation of **B** with **E**, and also the correlations of curls and divergences noted for static fields in the antiparallel system in connection with Eqs. (2–35) to (2–38).

In bringing this section to a close we briefly summarize the arguments on the two sides of the Sommerfeld-Kennelly unit question as we see them.

The Kennelly units are congenial to those who wish to use the traditional approach because (a) they are needed to give Eq. (2–43b) a form parallel to the electrostatic equation (2–42b); and because (b) the definition of **H** by $\mathbf{F}_m = \mathbf{H}'p_0$ [Eq. (2–23b)] would be spoiled if we introduced Sommerfeld units by substituting $\mu_0 p_0$ for p_0. (Our reasons for preferring $\mathbf{H}'p_0$ to $\mathbf{B}'p_t$ are indicated in Sec. 2.2.)

The Sommerfeld units, on the other hand, are congenial to those who use the antiparallel approach because (a) they are needed to give Eq. (2–44c) a form antiparallel to Eq. (2–42b); and because (b) they simplify the relation between magnetic moment and current. Under (b) we include the relation $\mathrm{curl}\mathbf{M} = \mathbf{J}_m$ for the volume-density of Amperian currents and the corresponding equation for the surface-density, as well as the equation for the magnetic moment of a linear circuit.

Possibly other considerations of importance have escaped us. In any case we hope that we have clarified the unit question by pointing out its intimate relation to teaching methods. Since our discussions lead to no clear mandate on the best pedagogical approach (see Sec. 2.7) we make no recommendation regarding magnetic units.

2.7 Magnetic Fields in the Classroom

Agreeing on the laws and correlations of Sec. 2.6, teachers will still disagree as to the best way of presenting them. Some will prefer to stick to the traditional approach, while others will go "all out" for the newer Amperian procedure. Both ways are right. Each has many variations, still right, and much room for the individual teacher's favorite demonstration experiments and other grace notes. The old principle, *de gustibus non disputandum*, comes near enough to this case to be a warning. Arguments here

should avoid dogmatic extremes, and they will be helpful only as they illuminate the problems.

For the traditional approach the chief positive argument has already been stated: It starts magnetic theory along familiar lines in conformity with the historical development of the subject and postpones the introduction of vector products as long as possible.

There are, however, a number of objections to this method that should be considered. Some teachers object to the emphasis that it places on a "fictional" concept (the magnetic pole) and to the fact that elementary textbooks using this approach are apt to give a false conception of the facts by creating the impression that every bar magnet has at its ends two point poles of fixed strength from which the magnetic field can be accurately computed according to Coulomb's law. Because the point pole is an idealization it may be argued that it is an illusion. In fact, in the minds of some the discrepancy between the point-pole approximation and the actual experimental situation seems to becloud the whole subject.[34]

In reply to these criticisms one can make two constructive suggestions for those who adhere to the traditional approach. The first suggestion is that the instructor make full use of the distinction between measured physical quantities and the idealized tools of physical theory[35] to point out that the crudity of the experiments which suggest a conceptual scheme in no way limits the precision with which the scheme itself is formulated. The theory of magnetic polarization, whether developed step-by-step from the pole concept, or from the notion of Amperian currents, is a well defined theory of great power and well able to account for the failure of the point-poles approximation at short distances. Its value is independent of the accuracy and range of the approximate formulas most helpful in enabling the beginner to grasp the initial concepts of the subject.

In the second place, it is clear that we need renewed emphasis on the necessity for treating the polar regions at the ends of a magnet as *distributions* of magnetic charge or magnetic pole-strength, which can be replaced by point poles *only* in a first crude approximation in order to simplify the computation of the field at large distances. The success of the concept of a distribution of magnetic pole-strength in explaining various experimental results, such as the properties of ellipsoidal magnets, long bar magnets, electromagnet pole pieces, etc., should be stressed. It would be helpful if tables of numerical data indicating the value and the limitations of the point pole approximation could find their way into elementary textbooks.[36]

For the Amperian approach the broadest argument is that it proceeds as directly as is practicable for a freshman toward the physical principles adequate for a satisfactory comprehension of the whole range of magnetic phenomena including induced electromotive forces. It unifies the treatment of the magnetic fields of bar magnets with those of electrical circuits and avoids the not-too-easy transition from one to the other that is inevitable in the magnetic-pole approach. It simplifies matters because it defines the magnetic field in free space at the beginning in terms of operations more suited to accurate measurement than the operation of measuring the force on a concentrated magnetic pole.[37]

Of course, the difficult spot in teaching by the Amperian approach lies in the initial step of defining the field by means of the transverse

[34] As one experienced teacher remarks, "No matter how carefully the statements about magnetic poles are qualified, the students who work formal problems involving magnetic poles almost invariably come to wrong conclusions. If later confronted with experimental tests of the departures from Coulomb's law at short distances, or with the fact that the attraction between unlike poles of a pair of similar magnets in contact is much greater than the repulsion of like poles in contact, they are at a complete loss to account for the observations and are likely to feel that the whole foundation of the subject of magnetism has evaporated."

[35] See Sec. 2.3.

[36] The committee feels the need of some new data—at least data more recent than the eighteenth century observations of Robison and Coulomb. A service would be rendered by any reader of this report who would undertake an accurate series of observations with carefully prepared magnets using the best modern materials. Incidentally, we should be interested to know just how effective the Robison design (with small iron spheres at the ends) actually is, in fixing the pole positions.

[37] An incidental practical advantage appears in the treatment of motors and generators. Every teacher has seen the confusion about the directions of the forces on the armatures of motors with the thumb and finger rules that have twelve geometric possibilities, of which six are right and six wrong. But because of the likeness of Amperian currents to the neighboring conduction currents, and because of the designs of modern motors (a.c. as well as d.c.) which place the field coils near the moving conductors, the directions of the forces can be found simply from the rule of attraction of like currents.

force on a suitably oriented current-carrying wire. A new physical concept, the magnetic field, and a new mathematical concept, the vector product, have to be introduced together. The average elementary student can well find himself baffled at this point even though the new vector **B** is defined in two steps, first fixing the direction in terms of lines of force revealed by iron filings and afterward fixing the magnitude by means of the mechanical force on a current element oriented at right angles to the line of force.

Advocates of the Amperian approach agree that the difficulty is a real one that can be met only by careful planning. Enough attention must be given to the lines, as lines, to make the student feel at home among them before any problem is raised about magnitudes. Iron filings should be used liberally. Also exploring coils, free to turn; also fairly straight, loose wires. There is plenty of work to do here before measuring any forces (see Sec. 3.34 of Outline II).

Only when he has this foundation is a freshman ready for the magnitude of **B**; but he can see it at this point in terms of either the force on a straight wire or the torque on an exploring coil. The latter view prepares him for the more usual method of measurement with the coil by electromotive impulse.

The arguments outlined here are not by any means all there are, but it is our belief that they are among the most important. To them we add a neutral suggestion: in using either approach don't omit good points suggested by the other. In the Amperian approach, for example, when you get around to quantitative comparisons of ferromagnetic materials, don't hesitate to use as the independent variable in the magnetization curve of a ring specimen the independently controllable number of ampere-turns per meter, alias **H**, even though historically the first reason for its use was a part of the other theory. In either approach don't hesitate to use either a scalar or a vector potential in free space, when you get around to needing it; nor to recognize that the external field of a solenoid can be calculated as if the solenoid was a bar magnet with poles, or *vice versa*. In short, enthusiasm for either system of logic is good; it enlivens the course; but it must be *for*, not against, anything that is true.

Finally, let us repeat: either approach is "right" and the student who goes on in physics must eventually learn both languages. The one thing we can say definitely here, about teaching as distinguished from subject matter, is that we members of the committee have profited greatly by our discussions of these facts and ideas, so we hope to share the profits with other physics teachers.

Radiation Pressure in a Sound Wave

ROBERT T. BEYER
Brown University, Providence, Rhode Island

THE concept of radiation pressure in a sound beam was first presented by Lord Rayleigh[1] nearly fifty years ago. Since that time the concept has frequently been invoked to justify experimental measurements on the absorption of sound in liquids.[2–4]

In a typical experimental set-up,[4] ultrasonic radiation is sent vertically upward in a container of liquid. A flat plate, or detector, is suspended into the liquid from one arm of a microbalance. The plane face of the detector is perpendicular to the direction of propagation of the radiation. When the source is emitting radiation, the apparent weight of the detector is less than it is in the absence of radiation. This loss in weight is attributed to the radiation pressure of the ultrasonic beam, which is, in turn, proportional to the acoustic intensity, so that the change in weight of the detector is also proportional to the intensity. A measurement of the change in weight

[1] Rayleigh, *Phil. Mag.* **3**, 338 (1902); **10**, 364 (1905).
[2] C. Sörensen, *Ann. d. Physik, Leipzig* (5) **27**, 70 (1936).
[3] Claeys, Errera, and Sack, *Trans. Faraday Soc.* **33**, 136 (1937).
[4] E. Hsu, *J. Acous. Soc. Am.* **17**, 127 (1945).

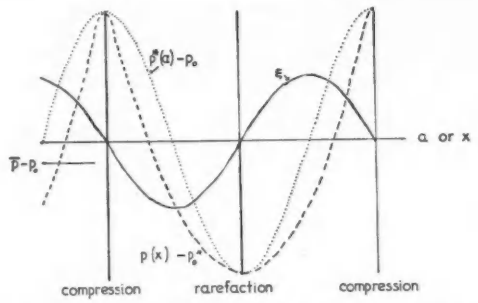


FIG. 1. Sound pressure as a function of position in a tube.

of the detector at two positions along the path of the beam makes possible the calculation of the sound absorption coefficient.

Considerable confusion has arisen, however, as to the precise meaning of the term radiation pressure, and as to what is actually measured by experiment.[5] During recent years, a number of papers have appeared in German journals,[6–8] clarifying the problem, while a lengthy mathematical formulation has been given by Brillouin.[9] However, to the writer's knowledge, no similar work has appeared in English. The purpose of this paper is to fill the gap, at least in part. The simple picture of radiation pressure which is given is largely that of Hertz and Mende.[6]

Rayleigh Radiation Pressure

Much of the difficulty arises from the confusing of two different definitions of the sound radiation pressure, those of Rayleigh and of Langevin. By the *Rayleigh radiation pressure* is meant the difference between the time average of the pressure at a point in a fluid through which a sound beam passes, and the pressure which would have existed in a fluid of the *same mean density*, at rest. The *Langevin radiation pressure*, on the other hand, is defined as the difference between the pressure at a wall and the pressure in the medium, at rest, behind the wall.

To evaluate the Rayleigh pressure, we consider a plane sound wave traveling in the positive x-direction and suppose that the fluid particle at the position $x=a$ when $t=0$ is displaced a distance ξ in the time t. Let the pressure at the point x be denoted by $p(x)$ and the pressure at the position of the particle, wherever it may be, by $p^*(a)$. This latter is the pressure at the instantaneous position of the particle whose initial position was at $x=a$. Thus $p^*(a)=p(a+\xi)$. (One could, of course, use x instead of a for this coordinate, but some confusion is avoided by the notation used.) Finally, let p_0 be the pressure in the fluid in the absence of sound.

The customary analysis[10] of plane wave propagation in a fluid leads to the equation

$$-(\partial p^*/\partial a)=(\partial^2\xi/\partial t^2). \tag{1}$$

To express the left side of this equation in terms of ξ, it is necessary to introduce an equation of state. For the usual case of a one-dimensional disturbance in an elastic fluid, Hooke's law may be written as

$$p^*-p_0=-(1/\beta)(V-V_0)/V \\ =-(1/\beta)(\partial\xi/\partial a), \tag{2}$$

where β is the compressibility. By differentiation of Eq. (2) and substitution of the result in Eq. (1), the well-known equation for a plane wave is obtained:

$$\rho_0(\partial^2\xi/\partial t^2)=(1/\beta)(\partial^2\xi/\partial a^2). \tag{3}$$

For a harmonic wave traveling in the positive a- (i.e., x-) direction, Eq. (3) has a solution of the form

$$\xi=\xi_0\sin(\omega t-kx), \tag{4}$$

where ξ_0 is the amplitude, ω the angular frequency, and $k=2\pi/\lambda$ the wave number. Hence

$$p^*=p_0+\frac{\xi_0 k}{\beta}\cos(\omega t-ka), \tag{5}$$

where p_0 is the hydrostatic pressure.

Since actual radiation pressure measurements are made by means of some detector placed in

[5] For example, Bergmann, *Ultrasonics* (English translation, Wiley, New York, 1938), p. 44. Calculations are given for the radiation pressure in liquids with the aid of Rayleigh's formula, which is valid only in gases. This error is corrected in the 3rd German edition (1942), but it is repeated in B. Carlin, *Ultrasonics* (McGraw-Hill, New York, 1949), p. 28.

[6] G. Hertz and H. Mende, *Zeits. für Physik* **114**, 354 (1939).

[7] C. Schaefer, *Ann. d. Physik, Leipzig* (5) **35**, 474 (1939).

[8] F. Bopp, *Ann. d. Physik, Leipzig* (5) **38**, 495 (1940).

[9] L. Brillouin, *Ann. de Physique, Paris* **4**, 528 (1925); *Rév. d'Acoustique* **5**, 99 (1936).

[10] See, for example, Morse, *Vibration and sound* (McGraw-Hill, New York, 1936), p. 187.

the path of the beam, it is important to determine the radiation pressure at a wall. Two cases may be distinguished:

(*a*) *Perfectly absorbing wall.* If the wall is perfectly absorbing, the same particle displacements are carried out in the wall as in the medium, so that the actual pressure on the wall is p^*. But from Eq. (5), $p_{Av}{}^* = p_0$ so that the Rayleigh radiation pressure at such a surface is equal to zero.

(*b*) *Perfectly reflecting wall.* In this case, standing waves are set up, with a node at the wall. Therefore, there is a pressure antinode at the wall. Hence

$$p^* = p_0 + (2\xi_0 k/\beta)\cos\omega t \qquad (6)$$

if we take $a=0$. Therefore, $p_{Av}{}^* = p_0$ as before.

Since we are dealing with linear equations, the case of a partially reflecting wall may be treated as a superposition of the two cases above. In any case the Rayleigh radiation pressure vanishes *for an elastic liquid*, under the proviso that the mean density of the liquid remains constant.

For an ideal gas, the equations for an adiabatic compression are

$$pV^\gamma = p_0 V_0{}^\gamma \qquad (7)$$

and

$$pV = RT,$$

where γ is the ratio of the specific heats, R the gas constant and T the absolute temperature.

If we consider the particles in a volume element $dadydz$, we have $p(x)=p^*(a)$ and $V/V_0 = 1+\partial\xi/\partial a$ so that

$$p^*(a) = p_0(1+(\partial\xi/\partial a))^{-\gamma}. \qquad (8)$$

If the waves are of small amplitude, we may expand the right side of Eq. (8), keeping terms of order $(\partial\xi/\partial a)^2$:

$$p^*(a) = p_0[1-\gamma(\partial\xi/\partial a) + (\gamma+1)(\gamma/2)(\partial\xi/\partial a)^2+\cdots]. \qquad (9)$$

Again, if a traveling harmonic wave passes through the medium with the form

$$\xi = \xi_0 \sin(\omega t - ka),$$

Eq. (9) becomes

$$p^*(a) = p_0 + p_0\gamma\xi_0 k\cos(\omega t-ka) + \tfrac{1}{2}p_0(\gamma+1)\gamma\xi_0{}^2k^2\cos^2(\omega t-ka),$$

and

$$p_{Av}{}^* = p_0 + \tfrac{1}{4}p_0\gamma(\gamma+1)\xi_0{}^2k^2. \qquad (10)$$

Now $\gamma p_0 = c^2\rho_0$, while $(\dot\xi^2)_{Av} = \frac{1}{2}\omega^2\xi_0{}^2$. Therefore,

$$\begin{aligned} p_{Av}{}^* - p_0 &= \tfrac{1}{2}\rho_0(\gamma+1)(\dot\xi^2)_{Av} \\ &= \tfrac{1}{2}(\gamma+1)E_{Av} \\ &= \tfrac{1}{2}(\gamma+1)I/c, \end{aligned} \qquad (11)$$

where $E_{Av} = \rho_0(\dot\xi^2)_{Av}$ is the mean energy density, I is the acoustic intensity, and c is the speed of the wave in the medium. The expression $p_{Av}{}^* - p_0$ is the Rayleigh radiation pressure, and the result given by Eq. (11) is often cited (incorrectly) as the radiation pressure for both liquids and gases.[11]

Langevin Radiation Pressure

As stated above, the *Langevin radiation pressure* is defined to be the difference between the pressure at an absorbing or reflecting wall and the pressure in the medium at rest behind the wall. In order to calculate this pressure, the relation of the pressure within and without the beam must first be determined. Hertz and Mende consider a plane sound wave, passing along a tube the cross section of which coincides with the cross section of the sound beam. The pressure is now to be calculated at a particular point A on the side wall of the tube. This pressure will be given either by the time average of the pressure at the point A, or the space average, over one wavelength, of the pressure at any instant of time. If the latter form is used, then the average pressure at A is given by

$$p_{Av} = \frac{1}{\lambda}\int_0^\lambda p(x)dx, \qquad (12)$$

where λ is the wavelength of the sound. This equation can also be expressed in terms of the

[11] It might be argued that the derivation for the case of an ideal gas has been carried one stage further in approximation than that for the elastic liquid, and hence that it should not be surprising that the former yields a non-vanishing value for the Rayleigh radiation pressure, while the latter does not. However, the second order term is an inherent part of the ideal gas approximation, while no such terms appear in the elastic liquid case. So long as one is considering small amplitude waves in elastic liquids and ideal gases, these results are equally correct. In any event, it will be shown later that the radiation pressure which is actually measured is independent of the equation of state of the medium.

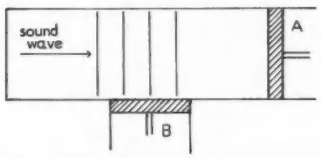


FIG. 2. Hypothetical arrangement to distinguish between Rayleigh and Langevin radiation pressures.

particles originally at a. Let

$$p^*(a)=p(x), \quad dx=da(1+\partial\xi/\partial a). \tag{13}$$

Then

$$p_{Av}=\frac{1}{\lambda}\int_0^\lambda p^*(a)(1+\partial\xi/\partial a)da.$$

For a traveling harmonic wave in an elastic liquid

$$p^*(a)=p_0+(\xi_0\omega/\beta c)\cos(\omega t-ka)$$

and

$$\partial\xi/\partial a=-(\xi_0\omega/c)\cos(\omega t-ka),$$

so that

$$\begin{aligned}p_{Av}&=(1/\lambda)\int_0^\lambda[p_0+(\xi_0\omega/\beta c)\cos(\omega t-ka)\\&\quad-(p_0\xi_0\omega/c)\cos(\omega t-ka)\\&\quad-(\xi_0^2\omega^2/\beta c^2)\cos^2(\omega t-ka)]da\\&=p_0-(\xi_0^2\omega^2/2\beta c^2)=p_0-((\dot\xi^2)_{Av}/\beta c^2),\end{aligned} \tag{14}$$

and since $c^2=1/\beta\rho_0$,

$$p_{Av}=p_0-\rho_0(\dot\xi^2)_{Av}=p_0-E_{Av}. \tag{15}$$

Thus, the average pressure on the walls of the tube is less than it would have been in the absence of sound. Hertz and Mende illustrate this with the diagram shown in Fig. 1. The solid curve represents the displacement of particles as a function of a (i.e., of their original position). The dotted curve gives the excess pressure $p^*(a)-p_0$ at an instant of time. Physically, this is the excess pressure at the point occupied by particles originally at a. In the vicinity of a compression (peak pressure), the length of the high pressure region is reduced. For this region, $\partial\xi/\partial a$ is negative so that $(1+\partial\xi/\partial a)da<da$. At a rarefaction, on the other hand, $(\partial\xi/\partial a)>0$, so that $(1+\partial\xi/\partial a)da>da$, and the pressure low extends over a greater region. This is shown in Fig. 1, in which the broken curve represents the variation of $p(x)-p_0$. The net result is a drop in the average value of $p(x)$ below p_0.

Now, suppose that a standing wave exists such that

$$\xi=\xi_0\sin\omega t\cos(\omega a/c), \tag{16}$$

where ξ_0 is the amplitude of the standing wave, and

$$\begin{aligned}p^*&=p_0-(1/\beta)(\partial\xi/\partial a)\\&=p_0+(\omega\xi_0/\beta c)\sin\omega t\sin(\omega a/c).\end{aligned} \tag{17}$$

The space average of p^* is then

$$\begin{aligned}\langle p^*(a)\rangle_{Av}&=(1/\lambda)\int_0^\lambda(p_0+(\omega\xi_0/\beta c)\sin\omega t\sin(\omega a/c))\\&\quad\times(1-(\xi_0\omega/c)\sin\omega t\sin(\omega a/c))da\\&=p_0+(1/\lambda)\int_0^\lambda(-\omega^2\xi_0^2/\beta c^2)\\&\quad\times\sin^2\omega t\sin^2(\omega a/c)da,\end{aligned} \tag{18}$$

or

$$\langle p^*(a)\rangle_{Av}=p_0-\sin^2(\omega t)(\omega^2\xi_0^2/2\beta c^2). \tag{19}$$

If the time average of the space average of p^* is also taken,

$$\langle p_{Av}\rangle=p_0-\omega^2\xi_0^2/2\beta c^2, \tag{20}$$

where the bracketed $\langle p_{Av}\rangle$ symbol indicates an average over *space and time*. Since $\beta c^2=1/\rho_0$ and $\langle(\dot\xi^2)_{Av}\rangle=\frac{1}{4}\omega^2\xi_0^2$,

$$\langle p_{Av}\rangle=p_0-\rho_0(\dot\xi^2)_{Av}=p_0-\langle E_{Av}\rangle, \tag{21}$$

where $\langle E_{Av}\rangle$ is the space and time average of the energy density.

The transition must still be made to the case of waves in an open fluid. The treatment here is again that of Hertz and Mende.[6] Suppose that a cylinder is placed within an elastic liquid of infinite extent and suppose that the cylinder is also filled with the liquid. We suppose two pistons are attached as shown in Fig. 2. In this case, a plane sound wave travels from left to right, its direction of propagation coinciding with the axis of the cylinder. We also suppose that the wave is absorbed at the piston wall A.

Case I. Let A be a movable piston, but let B be fixed. Let ρ_r and p_r be the density and pressure in the undisturbed medium outside the cylinder. From the foregoing analysis, it may be seen that the average pressure inside the cylinder

at A is also p_r, so that the movable piston is in equilibrium. The volume of the liquid remains constant (B being fixed), so that $\rho_i = \rho_r$, where ρ_i is the density of the liquid inside the cylinder. At point B the interior pressure is $p_r - E_{Av}$, so that a net pressure difference exists at the fixed piston. In this case, the net pressure at A is the Rayleigh radiation pressure and is zero in this particular case.

Case II. Suppose now that A is fixed and B is movable. Under these conditions, the piston at B will move into the cylinder, thus *compressing the liquid*, since the piston at A remains fixed. Equilibrium will be reached when the inside pressure at B is equal to p_r. However, since by Eq. (5) the inside pressure at A must be E_{Av} units larger than the inside pressure at B, it follows that the inside pressure at A is $p_r + E_{Av}$, and the net pressure at A is E_{Av}. This is the Langevin radiation pressure.

Finally, the movable piston B can now be replaced by direct communication with the liquid outside, so that the radiation pressure as measured in an extended medium will be the Langevin pressure.

Effects of Higher Order Terms

The derivations just given have applied only to the special cases of an elastic liquid and an ideal gas. More general expressions have been developed from which it can be shown that the value of Langevin radiation pressure is independent of the equation of state. The following is a brief summary of these expressions.

Rayleigh[1] has shown that if the instantaneous pressure at a point is given by $p = f(\rho)$ and is expanded in the form

$$p = f(\rho_0) + f'(\rho_0)\Delta\rho + \tfrac{1}{2}f''(\rho_0)(\Delta\rho)^2, \qquad (22)$$

where

$$f'(\rho_0) = c^2, \quad f(\rho_0) = p_0,$$

then the Rayleigh pressure p_R can always be expressed in the form

$$p_R = p_{Av}^* - p_0 = [1 + \rho_0 f''(\rho_0)/2c^2]E_{Av}. \qquad (23)$$

A simple calculation will show that this expression vanishes for an elastic liquid and reduces to Eq. (11) for an ideal gas.

In addition, Brillouin[9] has shown that for the same pressure-density relation (Eq. (22)), the mean pressure at a fixed point is given, for sinusoidal waves, by

$$\langle p_{Av} \rangle = p_0 + (\rho_0 f''(\rho_0)/2c^2)E_{Av}, \qquad (24)$$

where p_0 is the pressure in the fluid in the absence of sound. The Langevin pressure p_L is then given by

$$p_L = p_{Av}^* - \langle p_{Av} \rangle = p_0 + \left[1 + \frac{\rho_0 f''(\rho_0)}{2c^2}\right]E_{Av} - p_0 - \frac{\rho_0 f''(\rho_0)}{2c^2}E_{Av} = E_{Av}. \qquad (25)$$

Hence the Langevin radiation pressure, which is the quantity actually measured experimentally,[12] is independent of the equation of state of the medium, being equal to the mean energy density in all cases. It is worth mentioning that most experimental arrangements[2-4] do not measure the absolute Langevin radiation pressure, but only a quantity proportional to it (because of partial reflection at the detector). Also, the values of γ for gases and liquids are for the most part sufficiently close to one that the difference between $\frac{1}{2}(\gamma+1)E_{Av}$ and E_{Av} is scarcely greater than the experimental error involved in measuring them. It should also be noted that the Langevin value is of the same form as that for the radiation pressure due to a plane electromagnetic wave.[13]

[12] J. Hartmann and T. Mortensen, *Phil. Mag.* **39**, 377 (1948). In this paper experimental measurements were undertaken to find out which of the two radiation pressures was actually measured. The experimental results agreed with the Langevin form rather than the Rayleigh.

[13] See, for example, F. K. Richtmyer, *Introduction to modern physics* (McGraw-Hill, New York, 1928), p. 191.

It is a curious fact that, having minds tending to the infinite, with imaginations unlimited by time and space, the limits of our exact knowledge are very small indeed.—HENRY A. ROWLAND.

Exchange Forces

DAVID L. FALKOFF
University of Notre Dame, Notre Dame, Indiana

I. Introduction

UNDOUBTEDLY, one of the most challenging problems of contemporary physics is the determination of the nature of nuclear forces. In particular, the force between the "elementary" nucleons, the neutron and proton, is of central interest. For, since all nuclei are now regarded as compounded from these basic constituents, a knowledge of this force is indispensable for the understanding of even the simplest of nuclear phenomena.

The most striking evidence thus far obtained about this fundamental force has come from recent experiments with the newly built California synchrocyclotron, which accelerates deuterons to an energy of two hundred million electron volts (200 Mev). These experiments provide the first direct confirmation for the bold and unorthodox hypothesis, suggested some fifteen years earlier, that these forces are "exchange forces."

What is meant by *exchange forces?* How do these differ from the ordinary forces to which physicists are accustomed? What is the evidence which led to their prediction in nuclear phenomena long before any experimental test was practical? How can the idea of such forces be made plausible in terms of a reasonable mechanism? These are some of the questions which we now discuss.

II. Exchange Forces, Billiard Balls, and the Uncertainty Principle

The essential difference between exchange forces and ordinary forces can be effectively illustrated by a simple and familiar example from classical mechanics: the head-on collision between two smooth and perfectly elastic billiard balls.

Let us suppose the billiard balls are exactly the same except for their markings, one being all white, and the other white with a black figure eight on it—an unconventionally colored eight-ball. Now let the balls approach each other along the same line with oppositely directed velocities, as illustrated in Fig. 1(a). A straightforward application of the laws of conservation of energy and conservation of momentum then gives the motion of these balls after collision: namely, each ball recoils with the velocity which the other ball had before collision, as shown in Fig. 1(b).

Thus, the net effect of the collision has been for the two billiard balls to exchange momentum (and energy). This situation seems quite ordinary and therefore, in spite of the occurrence of the word *exchange* in the previous sentence, we shall refer to this collision as being governed by *ordinary* forces. More generally, any force occurring in classical mechanics will be called an ordinary force. Gravitational and electromagnetic forces are important examples of ordinary forces.

How would the results of the collision in Fig. 1 be changed if the force between billiard balls were not ordinary, but of the exchange type? Whatever exchange forces may be, one must, in the absence of any compelling evidence to the contrary, still require that the conservation laws of energy and momentum be satisfied. Is there any motion other than that shown in Fig. 1(b) which is consistent with these conservation laws? Clearly, an equally admissible motion after collision is one in which each ball continues with its original velocity unchanged, as shown in Fig. 2(b). But how can such a motion be possible? Surely the colliding balls cannot get through each other; this would violate one's sense of what a well-behaved billiard ball will do. An equally farfetched possibility, but at least one which will preserve the rigidity and impenetrability which one associates with billiard balls, is

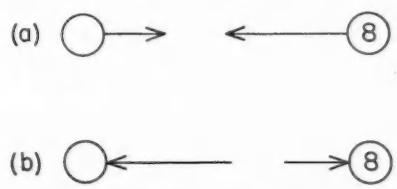


FIG. 1. Head-on collision between elastic billiard balls obeying ordinary forces; (a) before collision, (b) after collision.

to suppose that during collision, the figure eight on the eight-ball somehow got completely transferred from the eight-ball to the white ball. Quite irrespective of the mechanism which one may devise to account for the motion illustrated in Fig. 2, one can describe the over-all collision by saying it is due to an exchange force. In this way one implicitly defines an exchange force (acting between billiard balls) as *that force which yields for the resultant motion what one would get if one first obtained the (classical) motion using ordinary forces and then interchanged (or exchanged) both balls!*

This is indeed a strange kind of force and certainly it was not billiard ball experiments which suggested such anomalous forces. How were physicists led to consider such a possibility seriously? Suppose the billiard balls are taken to represent electrons; one can now think of

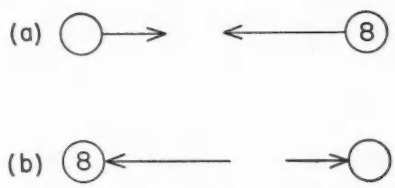


FIG. 2. Head-on collision between elastic billiard balls obeying exchange forces; (a) before collision, (b) after collision.

charged billiard balls. What will the figure eight on the eight-ball represent? Its original function was to enable us to distinguish between the two balls. But electrons are all alike. We must, therefore, discard the figure eight and consider instead two physically identical (white) billiard balls. Let us now look closely at Figs. 1 and 2 when we remove the number from the eight-ball. Do not both collisions for ordinary and exchange forces reduce to the same description as shown in Fig. 3? A sharp-eyed billiard player might object that they are not the same. For he can keep his eye on the ball and see whether or not the balls have changed places during the collision. Thus, even though the balls may be indistinguishable with respect to all physical properties such as mass, shape, color etc., nevertheless they can still be distinguished by their trajectories, i.e., by the distinct positions which they have

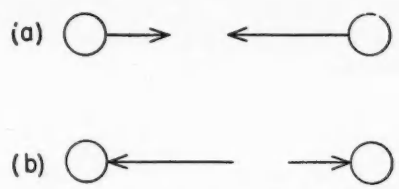


FIG. 3. Head-on collision of identical billiard balls; (a) before collision, (b) after collision with ordinary or exchange forces.

occupied and which we could have measured as precisely as we wished at all times.

It is here that the inadequacy of charged billiard balls as models for electrons is most forcibly evident. For the billiard balls obey classical mechanics, and classically it is quite legitimate to imagine an infinitely sharp-eyed billiard player who will never lose sight of the ball. But electrons are governed by the laws of quantum mechanics. The fundamental wave-particle duality of the electron in the quantum-mechanical description is most cogently formulated by means of the Heisenberg uncertainty principle. In the familiar form $\Delta p \Delta x \sim h$, it asserts that if, for example, one has exact knowledge of the momentum of an electron, then one must remain completely ignorant of its position. Hence one must abandon any hope of following the motion of an electron in detail. What is the consequence of this principle for our collision experiment in which we now regard the billiard balls as electrons? It means we cannot know (in Fig. 3) which electron it is that we detect coming out on the left, and which has emerged from collision on the right. A physical measurement of the outgoing electrons can at most yield the information that an electron comes out in each direction with some definite velocity. And this information (represented in Fig. 3) is, as we have seen, equally consistent with either the ordinary force (Fig. 1) or exchange force description (Fig. 2).

Thus from the quantum theory viewpoint, the use of the exchange force description of the interaction between identical particles appears just as natural as the ordinary force description, and accordingly both descriptions are treated on an equal footing in all applications of the theory. One might indeed assert that exchange forces are not only consistent with, but are in fact a neces-

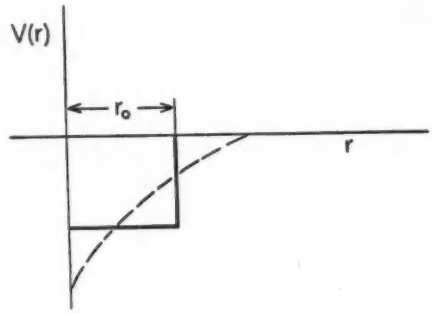


FIG. 4. Potential energy curves which can represent the short-range, strong, attractive neutron-proton force.

sary consequence of, the "operational" interpretation of the principles of quantum mechanics as applied to physically indistinguishable particles, and in this respect exchange forces represent no new hypothesis beyond what is already contained in quantum mechanics. The soundness of the theory as applied to atomic physics has been overwhelmingly established by its success in explaining numerous phenomena which were wholly beyond the reach of earlier theories. (One of the most notable successes was the development of a satisfactory theory of the chemical bond between atoms.)

To see how exchange forces made their way into nuclear physics, we forsake the billiard table, and devote the next section to a brief review of some basic background on nuclear forces.

III. On Nuclear Forces

The problem of nuclear forces has presented a very different kind of challenge to physicists than did the problem of understanding the atom earlier in the century. For in the case of the atom, Rutherford's classic experiments on the scattering of alpha-particles by atomic nuclei firmly established the present atomic model of a positively charged nucleus with the electrons far outside. And Bohr's bold and successful theory of the hydrogen atom which followed soon after made it clear that although the laws of classical mechanics must yield to some form of a new quantum mechanics, nevertheless, the law of force between electrons and nucleus must still be the classical Coulomb force. The subsequent revolutionary developments in quantum mechanics have not altered this basic fact. The laws of mechanics have had to be changed in atomic physics but not the law of force.

In nuclear physics the situation is just the reverse. Theoretical physicists have thus far proceeded with few qualms on the assumption that quantum mechanics is the correct mechanics for nuclear phenomena. But nothing in any other part of physics has suggested that any of the previously known forces (e.g., gravitational or electromagnetic) could adequately account for the tremendous energies encountered in nuclear physics. Accordingly, the existence of a wholly new force in nature, the nuclear force, had to be admitted and then investigated. In studying this force experiment has necessarily been the main guide, and theory has played a comparatively subordinate role. For with the force law unknown, there has been no classical analog or intuition on which to rely as was the case in the study of the atom with its classical force.

One of the most fruitful tools for exploring nuclei has been the type of experiment introduced by Rutherford in his pioneer work, namely, the scattering experiment in which one sends a collimated beam of monoenergetic particles at a target of known nuclei and observes the results of bombardment. There is now a vast literature on the results of such experiments with different nuclear projectiles (neutron, proton, deuteron, etc.,) of various energies and with almost all stable nuclei as targets. The analysis of these and other experiments, while far from yielding any force law as well-defined as the Coulomb law for atomic interactions, has definitely shown that the force law for the interaction between neutron and proton must have the following three properties:

1. The force must have an extremely *short range*, $r_0 \sim 2 \times 10^{-13}$ cm. This means that there is no detectable effect of a neutron on a proton unless they approach each other within this critical distance r_0—a behavior in marked contrast to the Coulomb force, the effects of which can be felt by a charged particle at any distance from any other charge, and for which, therefore, it is meaningless to speak of a definite range. It is interesting that scattering experiments

show also that the volumes of nuclei are given closely by $V \sim (4\pi/3)(r_0/2)^3 A$, where A, the mass number of the nucleus, is the total number of neutrons and protons constituting the nucleus. This shows that the nucleons are packed roughly a distance r_0 apart in any nucleus.

2. The neutron-proton force is a very *strong attractive* force. This is an immediate consequence of the fact that the deuteron, consisting of one neutron and proton, is a stable nucleus which requires an energy of 2.2 Mev to dissociate it.

3. The binding energies of nuclei are proportional to the mass number A. (The binding energy of a nucleus is defined as the energy corresponding to the difference between the total mass which the neutrons and protons constituting a nucleus would have when not combined in the nucleus and the actual mass of the nucleus, as measured, say, on a mass spectrograph. This energy difference represents the energy which must be supplied to dissociate completely all nucleons from the nucleus.) This characteristic of nuclear binding energies is called the *saturation property* of nuclear forces for the reason that it can be accounted for only if each nucleon in nucleus is assumed to interact with only a limited number of other nucleons. One can easily verify that if, on the contrary, each nucleon could interact equally with all other nucleons, then the binding energy should be at least proportional to $A(A-1)/2$, the number of distinct interacting pairs. Such a proportionality is not observed.

The first two properties of the nuclear force can be adequately represented by a simple square-well-shaped potential $V(r)$, where r is the distance between a pair of nucleons,[1] as shown in Fig. 4. The range is taken as $r_0 \sim 2\times 10^{-13}$ cm, and the depth is adjusted (to about 21 Mev) to fit the observed strength of the nuclear interaction. Thus far, for reasons which we shall presently see, experiments have not been sensitive enough to tell much about the shape of this potential. Another one of many equally satisfactory potentials is indicated by the dotted curve in Fig. 4.

Are such potentials representing a strong, short-range, attractive force consistent with the saturation property of nuclear forces? Clearly not, for if the force is purely attractive, then there is nothing to stop each nucleon from getting as close to any and all others in a nucleus as it pleases. But to explain the direct proportionality of both the binding energies and volumes of nuclei to the mass number A, one must assume there is a repulsive force which keeps the nucleons from coming too close together. Now the saturation properties observed in nuclei are very similar to the properties of ordinary bulk matter such as solids and liquids. In these cases, an interatomic potential of the type shown in Fig. 5, which is attractive down to an equilibrium distance d, and then sharply repulsive for closer distances between atoms, does account for the saturation behavior. For the repulsive force keeps the atoms a fixed distance d apart, so all atoms in a solid or liquid occupy equal volumes of the order d^3. This separation also restricts each atom to interact strongly only with its nearest neighbors, thus making the binding energy proportional to the number of atoms. Moreover, such potential energy curves were shown to be a consequence of quantum mechanics as applied to two atoms in which one assumed the ordinary Coulomb force between electrons and positively charged atomic nuclei, but in addition included the possibility of the exchange of the valence electrons in the interacting atoms.

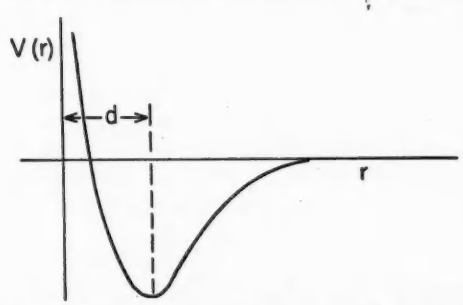


FIG. 5. An interatomic potential energy curve exhibiting both attraction and repulsion.

[1] Except for the specifically Coulomb part of the interaction between two protons, all the evidence so far is consistent with the convenient assumption that the purely nuclear force interaction is the same between neutron and neutron, neutron and proton, or proton and proton.

This remarkable fact, that a rather complicated force law which is in part repulsive and in part attractive could be equivalent to a simple, ordinary, purely attractive force together with an exchange of the interacting particles, is what led Heisenberg, as far back as 1932, to suggest that the nuclear force between neutron and proton was also of an exchange type.

Why is it that no direct experimental test of this hypothesis was made until the recent experiments at Berkeley? The answer to this question is connected with the uncertainty principle and the limitations imposed on measurements by the dual wave-corpuscular nature of matter. Just as the ultimate resolving power of an optical microscope is fixed by the finite wavelength of light, so also, in any experiment in which a nucleus is to be probed by a beam of incident particles, the information which one can glean is limited by the de Broglie wavelength of the incident particles. In particular, if one hopes to obtain detailed information about the shape of the interaction potential between neutron and proton by bombarding protons with neutrons, then one must require that the wavelength of the neutrons must be much smaller than the range of interaction r_0. (Quantum mechanically, a better yardstick than λ is the wavelength divided by 2π, or λ as it is written; for it is $\hbar$, Planck's constant h divided by 2π, which enters in the precise formulation of the uncertainty principle.) Using the de Broglie relation $\lambda = \hbar/mv = \hbar/(2ME)^{\frac{1}{2}}$, where E is the kinetic energy of the neutron and M its mass, and converting energy from ergs to million electron volts (Mev), the condition $\lambda \ll r_0$, is found to require $E \gg 5$ Mev. Actually, another factor of two in energy is required because half of the kinetic energy of the incident neutron will go into motion of the center of mass of the neutron-proton system and thus only half is available for the relative motion of neutron and proton. Since the energies available in laboratories until recently did not meet this requirement, it is clear why more conclusive experiments on the neutron-proton interaction were not feasible.

The beautiful and decisive high energy neutron-proton scattering experiments which have now been made using the 90-Mev neutron beam from the synchrocyclotron will not be further discussed here; their description and analysis merit separate consideration.[2] It is sufficient for our purposes to emphasize that they firmly establish the existence of an exchange force between neutron and proton. As a matter of fact, these experiments indicate that the nuclear force consists of about fifty percent exchange force, i.e., neutron and proton will exchange places in about half their encounters. How this can be interpreted with a plausible and useful model is the subject to which we next turn.

IV. Exchange Forces and Mesons

We have seen that exchange forces were first introduced into the theory of nuclear forces purely by analogy with electron-exchange effects in atomic interactions. Like many an analogy, the parallel has been suggestive of a fruitful point of view, but it also has its inconsistencies.

For electrons, the basis of the exchange interpretation was twofold: First, all electrons have identical physical properties and, second, the uncertainty principle made it impossible to follow a collision in sufficient detail to differentiate the particles without interfering with the experiment. Indeed, the exchange force interpretation could even be held to be a unique consequence of these two basic tenets of quantum mechanics. However, in considering the neutron-proton interaction, the first argument is no longer valid: neutron and proton do not have identical physical properties; they are easily distinguishable in virtue of their differing mass, charge, and magnetic moment. Thus the theoretical reasons for assuming an exchange interaction between neutron and proton were not at all compelling. But once theoretical physicists had been divested of all classical inhibitions about exchange forces, there was no difficulty in imagining many different kinds of possible exchange interactions, for example, interactions in which the positions of neutron and proton were interchanged, or only their spins, or both position and spin coordinates!

With theory in such an inconclusive state, a new idea or experimental discovery was needed. Theory came first. In 1935, the Japanese physicist, Yukawa, postulated the existence of a new elementary particle, the meson, to provide a

[2] Hadley, Kelly, Leith, Segrè, Wiegand, and York, *Physical Rev.* **75**, 351 (1949).

mechanism for the nuclear exchange force in the theory which he propounded. The attractiveness of this theory was considerably enhanced by the discovery of the predicted meson[3] two years later in cosmic-ray experiments, and the meson theory of nuclear forces, as it is now called, has been a major concern of theoreticians ever since.

The connection of mesons with nuclear forces can best be introduced (as indeed Yukawa originally did) by analogy with the quantum theory for the electromagnetic interaction of charged particles. This latter theory is not so formidable as it may sound, since after all Coulomb's law still gives the force between charges at rest in quantum mechanics as well as in classical mechanics. However, the way of looking at the static Coulomb field is quite different in the quantum theory. Instead of thinking of the force between charges as being due to lines of electric force continuously distributed in space, one can take the corpuscular view of the electromagnetic field and suppose that the force between charges is transmitted by photons. One then speaks of the electromagnetic field as being "quantized," the particles resulting from this quantization of the field being photons. More precisely, the interaction between two charges can be interpreted as due to the emission of photons by the one charge and subsequent absorption of these photons by the other. On carrying through the quantum-mechanical calculations based on this picture one arrives at all the familiar classical expressions.[4]

The above description must be qualified in one important respect. As it stands, it seems to predict, incorrectly, that one should observe radiation from charges at rest. To remove this objection, it is necessary to draw (the purely quantum-mechanical) distinction between *real* and *virtual* photon emission: Namely, a virtual photon emission (or absorption) can be characterized as one which cannot be detected because it would violate the energy conservation law. For example, an electron at rest cannot emit a photon of any frequency ν, having energy $h\nu$ and momentum $h\nu/c$, because it would be impossible for the electron to remain at rest and still maintain the constancy of energy and momentum for the system of electron plus photon. But it is convenient, in this case, to speak of virtual photon emission. On the other hand, if one were to supply energy to the electron by accelerating it, then this would make possible the *real* emission of photons which would be observable as the electromagnetic radiation from the accelerated charge. (This distinction between virtual and real emission is discussed again in more detail below.)

Let us now examine the force between neutron and proton and see how far analogy can lead. Just as the force between charges can be interpreted in terms of an electric field, so shall we now regard the neutron-proton force as due to a nuclear field, called the meson field. When this field is quantized, one obtains the particle associated with the field, and it will, of course, be called a meson. The meson is then assumed to transmit the force between nucleons by being (virtually) emitted and absorbed by the latter in the same way that the photon provided the mechanism for the interaction between electrically charged particles. In this way the theory aims to encompass both an explanation of nuclear forces and the properties of mesons in a unified manner.

If one identifies these mesons with the particles observed in cosmic-ray phenomena (and only recently produced in the laboratory), then in pursuing this analogy one must take into account the two essential differences between such mesons and photons. In the first place, mesons are found to have a finite (rest) mass in contrast to the zero (rest) mass of the photon. The mass of the π-meson, denoted here by μ, is of the order of three hundred times the mass of the electron. Second, mesons are observed to be electrically charged with charge $\pm\epsilon$, where ϵ is the charge of the electron, while photons are electrically neutral. These differences give rise to the fundamental dissimilarities between electromagnetic and nuclear forces.

The finite rest mass, μ, of the meson may be shown to imply that the meson field has a range of the order of $\hbar/\mu c$. Inserting the numerical values for $\hbar$, μ, and c, this is seen to yield a range for the meson field—and hence for nuclear

[3] At the present writing, there have been identified at least three distinct kinds of mesons differing in charge or mass, or both. The theory of their relationship to each other and nuclear forces is still in a very preliminary stage. To simplify the discussion we shall refer explicitly to only one kind of meson, the so-called π-meson. A good up-to-date account of recent meson experiments and their interpretation is given by J. M. Keller, *Am. J. Physics* **17**, 356 (1949).

[4] This may seem like a roundabout way to derive such a simple relation as Coulomb's law. But of course, the quantum theory of radiation does more than reaffirm classical theory where the latter is applicable; it can treat of much more that is completely beyond the scope of classical electromagnetic theory. For example, it accounts completely for spontaneous radiative transitions in atoms.

forces—of the order of 10^{-13} cm, which is in excellent agreement with the range as determined by nuclear scattering experiments. Indeed, in this way Yukawa predicted the mass of the meson from the observed range r_0.

The relation between the mass of the meson and the range of the meson field can be derived in a way which at the same time sheds light on the concept of virtual emission. We have earlier referred to a virtual emission as one which cannot be detected because it violates conservation of energy. One may wonder what point there is in speaking of a process which cannot be detected. To answer this, one is inevitably led to consider more closely the limitations on the measurement or detection of physical events by quantum mechanics as embodied in one or another formulation of the uncertainty principle. Here we use the form applicable to the conjugate variables, energy and time: $\Delta E \Delta t \geq \hbar$, which asserts that if the spread in time during which one has knowledge of a physical system is Δt, then the corresponding uncertainty in the energy of the system is $\Delta E \sim \hbar/\Delta t$. This has the consequence that physical processes (such as meson emission) which cannot be detected when the time of observation is long enough to insure a negligible uncertainty in the energy and whose occurrence would violate energy conservation, can nevertheless occur provided they do so for a sufficiently small time interval. For then the uncertainty in energy can save the energy balance for the brief duration of the process. Such transitory processes, i.e., ones which could be detected only in an extremely short time and only if sufficient additional energy is imparted to the system under observation by the measurement itself, are what are properly called "virtual."

We can now make quantitative use of this criterion as applied to virtual meson emission. The time interval Δt in which one could observe a meson emitted from a nucleon with velocity v is of the order $\Delta t \sim r/v$, where r is the extent of the distance from the nucleon at which the meson can be found, i.e., it is the range of the meson field. The corresponding uncertainty in the energy of the meson is $\Delta E \sim \hbar v/r$. How large must this uncertainty in energy be in order to guarantee conservation of energy for the virtual emission? It must be at least μc^2, the energy required to create the meson of mass μ. Thus

$$\Delta E \sim \hbar v/r \geq \mu c^2;$$

and since $v < c$, one gets $r < \hbar/\mu c$, which gives an upper limit to the range of the meson field.[5]

The second property of mesons, that they are electrically charged, leads directly to a simple mechanism for the exchange force between nucleons.

According to the meson field theory for nuclear forces, one can represent the meson field due to

[5] This argument is due to G. C. Wick, *Nature* **142**, 993 (1938).

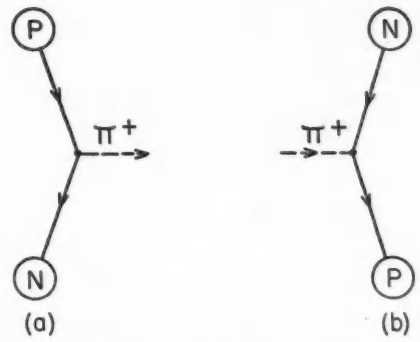


FIG. 6. (a) Transformation of a proton into a neutron by virtual emission of a positive meson: $P \rightarrow N + \pi^+$. (b) Transformation of a neutron into a proton by virtual absorption of a positive meson: $N + \pi^+ \rightarrow P$. Solid lines represent neutron or proton trajectories. Dotted lines represent virtual meson emission or absorption accordingly as the arrow is directed away from, or toward the common vertex.

a proton, say, by the (virtual) emission of positively charged mesons. Since electric charge must be conserved during any such emission the proton must be left electrically neutral after it has given its positive charge to the newly created meson. It was thus natural for theory to postulate that in this way the proton can be transformed into a neutron.[6] This transformation can be represented in the familiar notation for chemical reactions

$$P \rightleftarrows N + \pi^+,$$

where P denotes a proton, N a neutron, and π^+ a positive meson. The arrows point both ways to indicate that the inverse process is also possible, namely, a neutron can absorb a positive meson and thereby be transformed into a proton. These processes are shown schematically in Figs. 6(a) and 6(b), respectively. Equally feasible transformations involving the emission and absorption

[6] It is true that it is not charge alone which distinguishes neutron and proton. There are differences in mass and magnetic moment. However, it is one of the aims (not yet realized) of the meson field theory to account also for these differences by means of this same model. For example, the fact that the neutron has a magnetic moment, but no charge, is quite impossible to understand from either classical or quantum-electromagnetic theory, for how can there be any currents around to produce magnetic effects when their isn't even any charge? But in the meson theory, it is possible for each neutron to emit charged mesons, and one might hope to attribute the magnetic moment of the neutron to the charged meson field in the vicinity of the neutron.

of negative mesons according to the scheme

$$N \rightleftarrows P + \pi^-$$

are illustrated in Figs. 7(a) and 7(b).

We now apply these ideas to a collision between neutron and proton. We need only consider those collisions in which the nucleons come close enough together to be within each other's meson field. (For if this condition is not met they will not interact, each will continue with its original motion undisturbed, and then one really can't speak of a collision.) When the nucleons have come within their common force field we can describe their interaction in terms of meson emission and absorption in the following way:

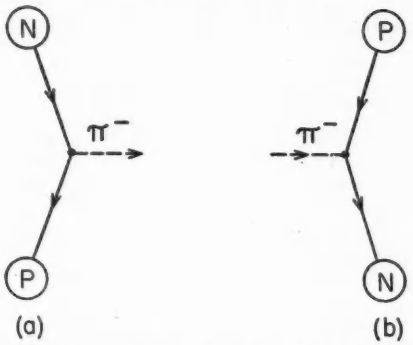


FIG. 7. Neutron-proton transformations involving negative mesons; (a) $N \rightarrow P + \pi^-$, (b) $P + \pi^- \rightarrow N$.

The incident neutron N can emit a negative meson (as in Fig. 7(a)) thereby becoming a proton which we here denote by P'. Then (as in Fig. 7(b)) the originally incident proton P can absorb this same negative meson and become a neutron N'. This meson transfer being consummated, the "newly born" proton P' and neutron N' continue to follow through with that motion, consistent with energy and momentum conservation, which N and P, respectively, were originally going to have after collision, as illustrated in Fig. 8. Thus, the net result of such a collision is that the incident neutron N has been transformed into a proton P' and the incident proton P has become a neutron N'. No meson need be detectable before or after collision,[7] the same negative[8] meson having been emitted and absorbed during the collision. Moreover, so far as the detection of the scattered nucleons is concerned, we see that there has been accomplished a *complete exchange of neutron and proton* in that everywhere that one would ordinarily expect to find the neutron N one finds instead the proton P' into which it was transformed; and similarly N' replaces P.[9] This is exactly what an exchange force is supposed to do according to the definition of Sec. II, given originally for billiard balls without reference to mesons, if one now applies it to the neutron-proton collision.

Indeed, if one specializes Fig. 8 to a head-on, neutron-proton collision, then (except for the meson exchange) it will look exactly like Fig. 2, if one labels the neutron as the white ball and the proton as the eight-ball. One sees that, of the farfetched mechanisms mentioned for the billiard ball exchange forces in Sec. II, the one which comes nearest to the nuclear mechanism is that one in which the figure eight was supposed to be transferred in some unspecified manner from the eight-ball to the white ball at the time of impact. The figure *eight* corresponds to the electric charge $\pm\epsilon$, which gets carried by the meson from proton to neutron, or *vice versa*. This kind of exchange force is often appropriately referred to as a *charge* exchange force.

As stated earlier, experiment now indicates that the nuclear force is a mixture of fifty percent

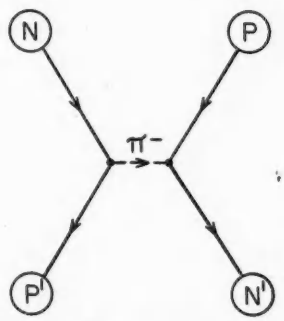


FIG. 8. Virtual emission and absorption of a negative meson during neutron-proton collision giving rise to charge exchange force.

[7] Whence comes the apt term "virtual" to describe these meson emissions and absorptions.

[8] A parallel process using positive mesons would give the same result.

[9] The distinction between N and N' (and P and P') which was made as an aid in describing the above mechanism for exchange is physically irrelevant. For what one can measure is *a* proton or *a* neutron coming out in a given direction, but the uncertainty principle prevents one from knowing *which* proton or *which* neutron.

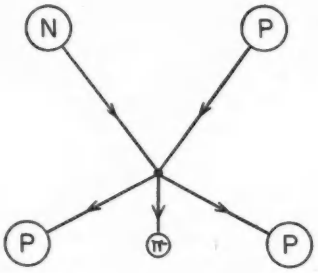


FIG. 9. Negative meson and two protons are outgoing particles in sufficiently energetic neutron-proton collision.

exchange force and fifty percent ordinary force. In terms of the meson field description this mixture can be simply interpreted as meaning that the meson, when virtually emitted by the neutron, has an equal chance to be (virtually) absorbed either by its parent neutron or by the proton during collision.

It may seem to some readers that the meson mechanism for exchange forces is rather *ad hoc* in that no mesons need be detectable either before or after a neutron-proton collision. The theory would surely be highly artificial if it only had use for the ephemeral virtual meson. But the theory also predicts that real observable mesons should be produced during a nuclear collision whenever the nuclei have sufficient kinetic energy to spare to create the meson and still satisfy energy conservation. Such a collision is illustrated in Fig. 9, in which neutron and proton collide and the neutron sheds a (real) negative meson becoming a proton. Thus the final products of this reaction are three particles: two protons and the negative meson. The total energy and momentum of the two incoming nucleons is distributed among the three outgoing particles, as is also their total electric charge.

Such meson production is very similar to the electromagnetic radiation (i.e., photons) produced when two electrons "collide," i.e., are accelerated and deflected due to the presence of each other's fields. A major difference is that to create mesons a certain minimum energy μc^2 must be available, whereas the creation of a photon is always possible as long as either charged particle has some kinetic energy, no matter how small. For a meson of mass three hundred times the electron mass this minimum rest energy is very nearly 150 Mev. If the experiment is such that the proton is initially at rest and the neutrons are the bombarding particles, then since half of the neutron's kinetic energy goes into the translational motion of the center of mass of the combined neutron-proton system, a minimum neutron kinetic energy of twice μc^2, i.e., very nearly 300 Mev, is necessary to create a meson in the neutron-proton collision. This minimum energy is called the *threshold energy* for meson production. In collisions between heavier nuclei such as alpha-particles with carbon nuclei, the threshold energy for the production of mesons is lower, since less of the kinetic energy of the bombarding nucleus is wasted on the motion of the center of mass of the system. Thus, real mesons have been produced with the California synchroton in the latter collisions,[10] but not in neutron-proton experiments, the energy of the neutron beam being far below threshold. For sufficiently higher kinetic energies, one should be prepared to expect the creation of more than one meson in a single nuclear collision. There is also the possibility that neutron or proton might completely disintegrate into mesons.

In conclusion, it should be pointed out that in spite of its great heuristic value and aesthetic appeal, the meson theory has not been very successful in a quantitative way. This is not the place to enumerate its shortcomings. Suffice it to say that it is the only theory currently available, but already, as in the many different kinds of mesons observed, it has been far outpaced by experiment. Further progress in theory awaits the many more new discoveries which are sure to be unfolded as experiment leads the way with the controlled laboratory investigations of high energy phenomena. Indeed, the exchange force experiments, while already a milestone, have (in more ways than one) barely brought us to the threshold of what promises to be a most exciting and productive era in physics.

The author would like to express his appreciation to his colleague, Professor C. Mullin, for a critical reading of this article during its preparation.

[10] E. Gardner and C. M. G. Lattes, *Science* **107**, 270 (1948); J. Burfening, E. Gardner, and C. M. G. Lattes, *Physical Rev.* **75**, 382 (1949).

Confetti and Turbulence*

EARL W. THOMSON
U. S. Naval Academy, Annapolis, Maryland

WHEN the physics teacher comes to the subject of fluid dynamics, no matter how elementary his presentation, he should introduce and show experimentally the concepts of streamlined flow, turbulence and vorticity. The mathematics of this portion of physics is so advanced that it is usually not presented beyond the development of Bernoulli's principle, the conservation of energy along a streamline. Even the idea that the pressure decreases as the velocity increases along a streamline seems contrary to *common sense* and requires some excellent instruction.

In both the lecture and the laboratory, the subject has too often been neglected experimentally because of the lack of easily operated apparatus. From fundamental ideas, demonstrated with an aspirator, a spool-card-and-pin, and perhaps, a restricted flow apparatus with manometer, it is a very big jump to wind tunnels. Several intermediate steps could profitably be used to make the transition smooth and memorable—water flow through a weir with streamlines shown by dye, or air flow with smoke as the indicator.

The apparatus employed in this demonstration uses a 16-inch electric fan as a source of air-flow with confetti as the indicator of turbulent areas. Several obstacles were constructed, each about 6 inches in the long dimension and 8 inches high, being of constant cross section in the horizontal plane. Among these solid figures were a flat plate, a rectangular box, a cylinder, a faired *teardrop* airfoil of fineness ratio 4, an airfoil with the lower side plane, and an airfoil with the lower side concave.

Each of these was placed in turn upon a dowel pin in the horizontal plate (about 26 in. $\times$30 in.); confetti was sprinkled uniformly over the plate, and the fan turned on to full speed from a distance of about six feet. The confetti is immediately driven out of the streamline and

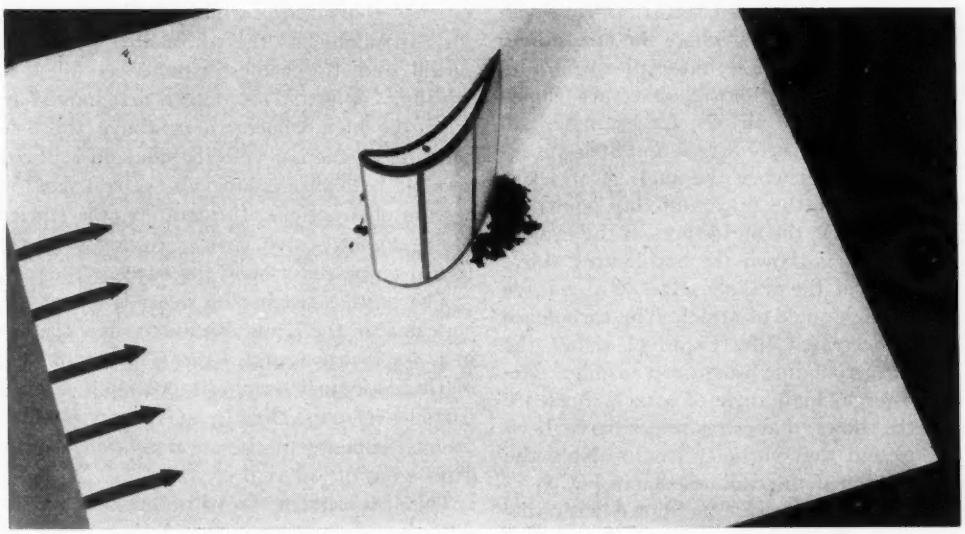


FIG. 1. Turbulence above the rear half of the trailing edge of a concave airfoil at high angle of attack. The arrows indicate direction of air-flow. (Official U. S. Navy Photograph.)

* Presented by title before the District of Columbia and Environs Section of the American Association for Physics Teachers, Goucher College, Baltimore, Maryland, 2 April 1949. Shown at Physics Colloquium, State University of Iowa, Iowa City, Iowa, 16–18 June 1949.

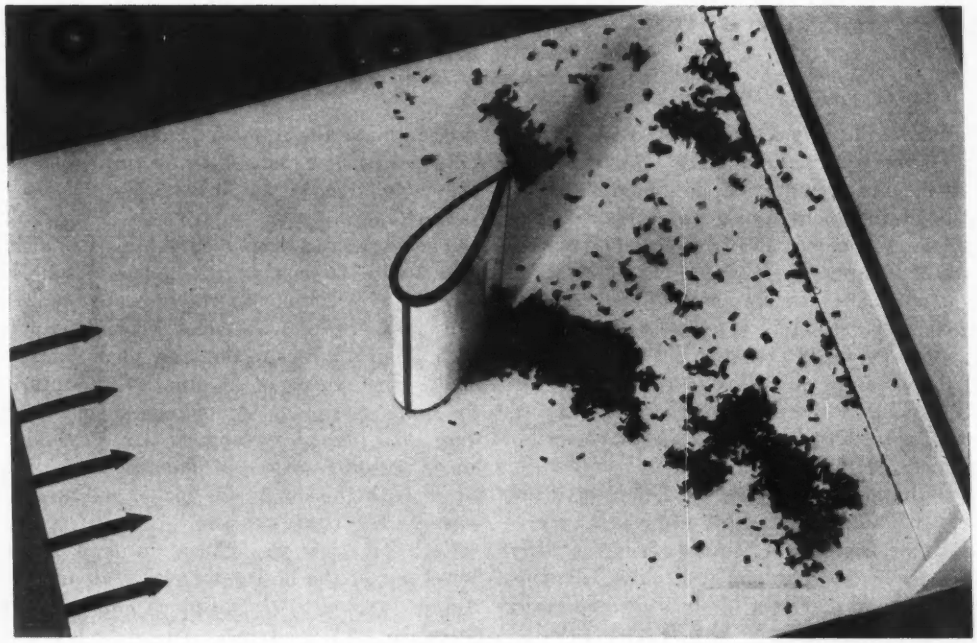


FIG. 2. Turbulence surrounding teardrop airfoil of fineness ratio 4:1 at high angle of attack, when backed by "confetti catcher." Note the high turbulence between the airfoil and rearward obstruction. (Official U. S. Navy Photograph.)

high velocity areas, leaving clean areas. However, in the low velocity areas, where the streamlines are broken up by vortices, the confetti piles up and the turbulent or *burble* areas are shown quite strikingly. The airfoils, for example, can be put at small angles of attack and little turbulence is shown, but when the angle of attack is increased beyond the maximum lift point the confetti piles up in the burble area at the trailing edge. In Fig. 1 is shown the burble area above the rear half of the trailing edge of a concave airfoil at a high angle of attack. The turbulence results in decreased lift. Figure 2 shows the turbulence surrounding a teardrop airfoil of fineness ratio 4:1 at high angle of attack. Presence of confetti shows that streamlines have been broken up and that vorticity results. Note the high turbulence at the trailing edge in Fig. 2.

Many other effects can be shown with this open low velocity *wind tunnel* in illustrating streamlining and turbulence. Because of the professorial fear of janitors we built a *confetti-catcher* and placed it about a foot to the rear of the trailing edge of the airfoils. Figure 2 shows entirely too well the whirling vortices set up by an obstruction, as the whole area between the airfoil and the confetti-catcher is filled with whirling confetti. The confetti first moved away from the high velocity area above the airfoil, piled up on the low velocity side and in the centers of the eddies created in the restricted area by the obstructions. Incidentally, the efficiency of the confetti-catcher was zero, for all the confetti disappeared around the edges.

The physics teacher for years has been using cork dust in the Kundt's tube to show the nodes in a stationary sound wave. The accumulation of the cork dust occurs at points of minimum particle velocity, that is, at the nodes. In the confetti experiment the accumulation also indicates areas of low velocity.

This experiment is of interest because it shows some of the concepts of fluid dynamics in a qualitative manner. Cork dust, dry sawdust, or even talcum powder can be used in place of confetti. Confetti can be used as an adjunct to science as well as a prelude to marital bliss.

On the Changing Status of Mesons*

JAMES N. SNYDER
University of Illinois, Urbana, Illinois

RECENT developments in meson theory have done much to bring some measure of qualitative order into a complex and baffling subject. The object of this paper is to point out historically several of the most important discoveries leading to the present knowledge of mesons, and to indicate how meson theories attempt to correlate these facts.[1]

I. Cosmic Rays and the Discovery of Mesons

One of the most striking properties of cosmic rays is their ability to penetrate great thicknesses of matter. During the 1920's many experimenters noted this fact; the experiments of Millikan and Cameron[2] on the penetration of the ionizing cosmic rays into mountain lakes are well known. These rays have been observed underground at depths as great as 1400 m water-equivalent and they probably penetrate much deeper.

This penetrating power of the cosmic-ray particles is quantitatively described by means of an absorption curve.[3] If N, the number of particles per unit time capable of penetrating a thickness of lead t, is plotted against t, a curve having the rough features of that shown in Fig. 1 is obtained. Such a behavior can be understood by assuming that cosmic rays consist of two types of particles, the soft component and the hard component. As the thickness of the absorber is gradually increased, the soft, or less penetrating component, is strongly absorbed, accounting for the steeper initial slope. By the time the knee of the curve is reached, most of these particles have been stopped. Further increase of the absorber thickness does not produce as strong a decrease of the more penetrating or hard component. It is the penetrating particles that are observed in mines and under great thicknesses of matter, which they are capable of penetrating with small energy loss.

The soft component has been shown to consist of high energy electrons, positrons, and γ-rays. Groups of many of these particles are often found associated with each other in a shower, and their presence in the cosmic radiation can be accounted for by the cascade theory[4] of cosmic-ray showers. By *Bremsstrahlung* an electron or positron can produce a γ-ray which, if energetic enough (energy >1 Mev), can in turn produce an electron-positron pair by the process of materialization. The resultant particles can then continue the process repeatedly, thus producing a cascade shower. Theoretically, the shower could be initiated by either an electron (positron) or a γ-ray. The nature of the initiating agent and its origin have not yet been decided by the existent data.

It may be mentioned for completeness that the cosmic radiation also contains showers of penetrating particles, nucleons (neutrons and protons), and, as was recently discovered,[5] heavier nuclei of atomic number 4 to 40.

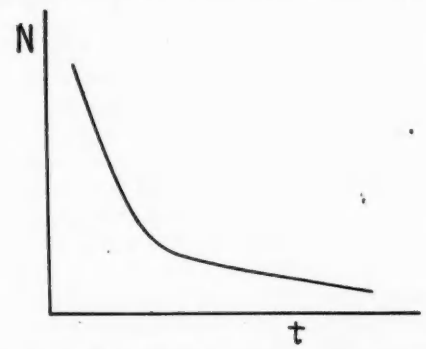


FIG. 1. A typical cosmic-ray absorption curve. A plot of the number of particles per unit time penetrating an absorber against the thickness of the absorber.

* A summary of an address to a Seminar on Advanced Modern Physics, Illinois Section of the American Association of Physics Teachers at the University of Illinois, Urbana, Illinois, June 1949.

[1] Excellent general references for the subject are: L. Janossy, *Cosmic rays* (Oxford Press, 1948); B. Rossi, *Rev. Mod. Physics* **20**, 537 (1948) and **21**, 1–183 (1949). See also Ref. 14.

[2] R. A. Millikan and G. H. Cameron, *Physical Rev.* **31**, 921 (1928) and **37**, 235 (1931).

[3] B. Rossi, *Zeits. f. Physik* **62**, 151 (1933); Street, Woodward, and Stevenson, *Physical Rev.* **47**, 891 (1935).

[4] H. J. Bhabha and W. Heitler, *Proc. Roy. Soc.* **A159**, 432 (1937); J. F. Carlson and J. R. Oppenheimer, *Physical Rev.* **51**, 220 (1937).

[5] Freier, Lofgren, Ney, Oppenheimer, Bradt, and Peters, *Physical Rev.* **74**, 213 (1948); *Rev. Mod. Physics* **21**, 101 (1949).

In this paper we are concerned with the hard component whose nature remained baffling for some time. Such particles of very high energy ($\sim 10^9$ electron-volts) were observed to pass through large thicknesses of matter with losses of energy which were very much smaller than those predicted for an electron by the Bethe-Heitler[6] theory. On the other hand, the absorption of the soft component conformed quite well to the theory. To explain the existence of the hard component, two alternate hypotheses were proposed:

(1) The particles were extremely high energy electrons and the Bethe-Heitler theory was invalid for such electrons.

(2) The Bethe-Heitler theory was valid at all energies and the particles were not electrons, but new and previously unidentified entities.

For some time hypothesis (1) had the wider credence, since there was also theoretical evidence for the inapplicability of electron theory at energies greater than 137 mc^2 (m is the mass of the electron and c is the speed of light).

However, in 1934 Williams[7] showed that, by means of a Lorentz transformation, a coordinate system could always be found with respect to which the electron could be regarded as moving slowly; i.e., if the Bethe-Heitler theory is valid at any energy (and it certainly is for small electron energies), it is valid at all energies. In addition, by 1937–38 improved experimental techniques enabled Neddermeyer and Anderson[8] to measure the mass of the particles comprising the hard component. Their results corroborated hypothesis (2), proving that the hard component consisted of particles which were called *mesotrons* or *mesons* of mass approximately 200 times the electron mass and of either positive or negative charge equal in magnitude to the electronic charge. Many other observers verified these measurements, and mesons became firmly established experimentally and formed the subject of many investigations.

One of the earliest phenomena observed in these studies was the anomalous absorption of mesons by the atmosphere. If the number of mesons arriving per unit time at sea level is compared with the number of mesons arriving per unit time under an absorber on a mountain top where the mass per unit area of the absorber is equal to the mass per unit area of the air between the mountain top and sea level, then these two rates were expected to be equal since both groups of mesons penetrated the same amount of material and should have been absorbed equally. However, various experimenters[9] showed that this was not the case; fewer mesons arrived at sea level. The implication is that mesons are lost, not only by absorption in matter but also in some other manner. The data given by these experiments could be explained by assuming that the meson was unstable, spontaneously decaying with a lifetime of the order of a microsecond and being lost. Coincidence methods[10] verified this assumption and, in addition, cloud-chamber results provided direct photographic evidence[11] that the meson was unstable. The only identifiable product of the decay was an electron. However, conservation of both energy and momentum in such a process requires that the meson decay into at least two particles. The second unobserved particle was assumed to be either a neutral meson or a neutrino, the latter having been introduced by Fermi[12] to serve a similar function in the theory of beta-decay.

II. The Origin of Meson Theory

During the years 1935–38, prior to the experimental discovery of the meson, Yukawa[13] and his co-workers in Japan developed a theory that in some ways anticipated the discovery of

[6] H. Bethe and W. Heitler, *Proc. Roy. Soc.* **A146**, 83 (1934).

[7] E. J. Williams, *Physical Rev.* **45**, 729 (1934).

[8] S. H. Neddermeyer and C. D. Anderson, *Physical Rev.* **51**, 884 (1937) and **54**, 88 (1938).

[9] D. H. Follet and J. D. Crawshaw, *Proc. Roy. Soc.* **A155**, 546 (1936); Auger, Ehrenfest, Freon, and Fournier, *Compt. Rend.* **204**, 257 (1937); B. Rossi, *Rev. Mod. Physics* **11**, 291 (1939).

[10] F. Rasetti, *Physical Rev.* **60**, 198 (1941); B. Rossi and N. Nereson, *Physical Rev.* **62**, 417 (1942) and **64**, 199 (1943).

[11] E. J. Williams and G. E. Roberts, *Nature* **145**, 102 (1940); E. J. Williams and G. R. Evans, *Nature* **145**, 818 (1940).

[12] E. Fermi, *Zeits. f. Physik* **88**, 161 (1934) and *Physical Rev.* **48**, 570 (1935); H. A. Bethe and R. F. Bacher, *Rev. Mod. Physics* **8**, 82 (1936).

[13] H. Yukawa, *Proc. Phys. Math. Soc. Jap.* **17**, 48 (1935); H. Yukawa and S. Sakata, *Proc. Phys. Math. Soc. Jap.* **19**, 1084 (1937); Yukawa, Taketani, and Sakata, *Proc. Phys. Math. Soc. Jap.* **20**, 319 (1938); and Yukawa, Sakata, Kobayasi, and Taketani, *Proc. Phys. Math. Soc. Jap.* **20**, 720 (1938).

mesons. Yukawa was attempting to account for the forces between nucleons in a way analogous to the field-theoretical treatment of the forces between electrons. According[14] to this latter treatment, electrons are regarded as exerting forces on each other in virtue of the fact that they are both coupled to the electromagnetic field. Under such circumstances, the coupling constant, i.e., the quantity which determines the strength of the interaction between the electrons and the electromagnetic field, is the charge of the electron e. The interaction can be imagined to take place as follows: one electron virtually emits a photon and thereby has its momentum changed; the photon is then absorbed by the second electron which has its momentum changed, i.e.,

$$e_1+e_2\rightarrow e_1'+e_2+h\nu\rightarrow e_1'+e_2', \qquad (1)$$

where the primes indicate that the particles have experienced momentum changes. By the elementary definition, this mutual change of momentum is equivalent to stating that the two electrons have exerted forces on each other.

To account for nuclear forces was quite a different problem. Nuclear forces are known to have a finite range outside of which they are negligible, and to be of an exchange nature. For example, the force between a neutron and a proton involves the repeated exchange of the charge between the two particles. Yukawa and co-workers investigated what kind of forces could arise between two nucleons which were respectively coupled to an intermediate field and what the properties of this field would be. The electronic forces could be accounted for by virtue of an intermediate field of quanta or photons which satisfied Maxwell's equations. So a set of modified Maxwell's equations was set up to describe the intermediate field. It was found that this procedure would yield a potential between nucleons given by the equation

$$V=g^2(e^{-r/R})/r, \qquad (2)$$

where r is the internucleonic distance, g is a constant which determines the strength of the coupling between the nucleons and the field, and R is a constant which can be interpreted as the range of the forces. This field is very different from the electromagnetic field responsible for electron interaction, for a finite rest mass must be associated with the *quanta* or *photons* associated with it. These were called "heavy quanta" by Yukawa; their mass μ is associated with the range of the forces by

$$R=h/\mu c, \qquad (3)$$

where $h=h/2\pi$, h = Planck's constant, and c = velocity of light. In order to account for the observed range of nuclear forces, it is necessary that μ be about 300 electron masses. If μ be placed equal to zero, Eq. (3) gives an infinite range of force, and Eq. (2) reduces to the well-known $1/r$ potential between electrons. In this case the equations of the field reduce to the Maxwell equations for the electromagnetic field which describes the photons of zero mass by virtue of which the electrons interact.

As they arise from the theory, these "heavy quanta" can be either neutral, positive, or negative. The magnitude of their charge is the same as that of the electron or proton. In these three cases the interaction between a neutron and a proton, for example, can be envisaged as

$$\begin{aligned} &N+P\rightarrow\mu^0+N'+P\rightarrow N'+P' \\ &N+P\rightarrow N+P'+\mu^0\rightarrow N'+P' \end{aligned} \qquad (4)$$

if the exchanged "quantum" is neutral, or as

$$\begin{aligned} &N+P\rightarrow\mu^-+P'+P\rightarrow P'+N' \\ &N+P\rightarrow N+N'+\mu^+\rightarrow P'+N' \end{aligned} \qquad (5)$$

if the exchanged "quantum" is charged. It will be noticed that the reactions (5) have the aspects of an *exchange force* mentioned above.

Yukawa also proposed to account for the beta-decay of nuclei by means of these heavy quanta. Previously beta-decay had been attributed[12] to a neutron in the nucleus spontaneously changing into a proton with the emission of an electron and a neutrino which was introduced to account for the continuous beta-spectrum

$$N\rightarrow P+e+\nu. \qquad (6)$$

By assuming that the heavy quantum was unstable and decayed into an electron and a neutrino, Yukawa accounted for beta-decay as the

[14] See, for example, G. Wentzel, *Einführung in die Quantentheorie der Wellenfelder* (Deuticke, Wien, 1943), an excellent reference on the field-theoretical treatments of mesons, electrons, and photons.

emission of a negative quantum by a neutron, which changed thereby into a proton, and the subsequent decay of the quantum into an electron and a neutrino

$$N \rightarrow P + \mu^- \rightarrow P + e + \nu. \qquad (7)$$

It remained to assign intrinsic angular momenta or spins to these new particles on the basis of the above reactions. It had long been known that the spins of the neutron, proton, electron, and positron were $\frac{1}{2}$ where the unit is $\hbar$. On the basis of reaction (6) the spin of the neutrino must also be $\frac{1}{2}$ since spins must add by the quantum-mechanical vector rule familiar from atomic line spectra. It is evident that with this rule the reactions (4), (5), and (7) are consistent with the heavy quantum, having a spin of either 0 or 1. The theories and properties of particles of spin 0 (scalar particles) and of spin 1 (vector particles) were thoroughly investigated by Yukawa and others.

It was only after these proposals had been made that the meson was discovered by Neddermeyer and Anderson in 1938. Yukawa then suggested that the meson and the "heavy quantum" were one and the same particle. The suggestion was universally adopted and for a decade proved very fruitful.

III. The Origin of Mesons

The meson, which is unstable, cannot possibly be the primary cosmic-ray particle, since it could not exist long enough to reach the earth from its origin in space. The neutron is also unstable, having a lifetime of 10–30 min., and there is good evidence[1] that electrons are not present in the primary radiation. By elimination this leaves the proton, and the assumption that the primary component of the cosmic radiation arriving at the upper atmosphere from exterior space consists of protons is reasonably well borne out by the latitude, and East-West effects.[1]

How then is the presence of mesons to be accounted for? An answer is provided by a field-theoretical treatment which has an analogy in electron theory. In an energetic collision between electrons, quanta (γ-rays called *Bremsstrahlung*) can be emitted in virtue of the coupling between the electrons and the electromagnetic field. Mesons are supposed to play the same role in relation to nucleons and nuclear forces that light quanta or photons play in relation to electrons and the forces between them. By analogy, mesons can be regarded as being produced by energetic nucleon collisions. In accordance with this picture the observed mesons are envisaged as arising in the collision of the extremely energetic primary protons with nucleons in the upper atmosphere. Such theories have been investigated by Heitler[15] and others with some qualitative success.

The scheme referred to below as "the old theory" and outlined in Sections I, II, and III was a fairly successful one. It accounted simultaneously for some of the qualitative aspects of meson production, nuclear forces, meson decay, and beta-decay, and was given wide credence as late as 1947. However, many experimental discoveries, a large number of them within the last two years, give evidence which this scheme is incapable of explaining. Hence it has been necessary to abandon it in favor of a new picture which will be discussed after a brief resume of the experiments which have made this step necessary.

IV. New Discoveries. Difficulties of the Old Theory

One objection to the above scheme is that it is unable to account quantitatively for some of the most important properties of nuclei and nuclear forces. However, it was not clear whether the difficulty lay in the proposed scheme or in the general unsatisfactory nature of the quantum-mechanical methods of calculation used. Very recent investigations of field-theoretical methods seem to point to the latter reason. Completely satisfactory and unambiguous methods of treating the meson field have not yet been developed.

Apart from theoretical objections, experimental evidence contradicting the theory was obtained. Experiments[16] on the scattering of

[15] W. Heitler and H. W. Peng, *Proc. Roy. Irish Acad.* **49**, 101 (1944); J. Hamilton, W. Heitler, and H. W. Peng, *Physical Rev.* **64**, 78 (1943); W. Heitler and P. Walsh, *Rev. Mod. Physics* **17**, 252 (1945); W. Heitler, *Rev. Mod. Physics* **21**, 113 (1949); and Lewis, Oppenheimer, and Wouthuysen, *Physical Rev.* **73**, 127 (1948).

[16] P. Ehrenfest, *Compt. Rend.* **207**, 573 (1938); and J. G. Wilson, *Proc. Roy. Soc.* **A166**, 482 (1938) and **A172**, 517 (1939); and F. L. Code, *Physical Rev.* **59**, 229 (1941).

sea-level mesons by nuclei in matter yield evidence that there is no appreciable coupling between mesons and nucleons that cannot be almost completely accounted for by the simple electromagnetic interaction which arises because both particles are charged. This definitely contradicts the hypothesis that mesons are produced in nucleonic collisions in the atmosphere.[17] To account for the observed flux of mesons as being produced by collisions of the inferred number of primary protons would require a strong coupling or interaction between nucleons and mesons. This strong coupling would in turn imply that mesons would be strongly scattered in collisions with nuclei. Observations show that this is not true.

The absorption of mesons by matter also yielded facts that were inconsistent with the old theory. Experiments on the decay of mesons at sea level stopped in various materials were carried out[18] by several observers with interesting results. Although the decay lifetime of both positive and negative free mesons is the same (about 2×10^{-6} sec), negative mesons decaying in matter have their lifetimes substantially shortened while the lifetime of positive mesons is unaltered. This shortening of the lifetime of the negative mesons is more pronounced in absorbers of higher atomic number. These results are interpreted as being due to a capture[19] of negative mesons by the positively charged nuclei of the material—this capture competing with the natural decay of the meson. The competition is naturally more in favor of capture for nuclei of greater charge. It is, in fact, observed that in the light elements the mesons predominantly decay; in the heavy elements they are predominantly captured. The two effects are approximately equal for the element of atomic number 12. The fact that mesons, which have a relatively long lifetime, can decay at all in condensed matter with the subsequent observation of the decay electron bears out the previous statement that mesons interact weakly if at all with nuclei.

[17] L. Nordheim and M. Hebb, *Physical Rev.* **56**, 494 (1939); V. F. Weisskopf, *Physical Rev.* **72**, 510 (1947).

[18] Conversi, Pancini, and Piccioni, *Physical Rev.* **71**, 209 (1947); H. K. Ticho, *Physical Rev.* **72**, 255 (1947); H. K. Ticho and M. Schein, *Physical Rev.* **73**, 81 (1948); and G. E. Valley and B. Rossi, *Physical Rev.* **73**, 177 (1948).

[19] J. A. Wheeler, *Rev. Mod. Physics* **21**, 133 (1949).

The explanation of the capture process presents great difficulties. On the basis of the old theory this process must be regarded as the absorption of the meson by a nuclear proton which is transformed thereby into a neutron

$$\begin{array}{ccccc} P & + & \mu^- & \rightarrow & N, \\ \frac{1}{2} & & 0 \text{ or } 1 & & \frac{1}{2}, \end{array} \qquad (8)$$

where the spin assignments of the various particles are explicitly indicated. However, the rest mass of the meson represents an energy of 100 Mev which is then available to excite the nucleus. Such an excited nucleus should evaporate with the emission of many energetic particles. No such explosions have been observed when mesons are captured at sea level.

On the other hand, the capture of negative mesons at high altitudes produces entirely different results.[20] Negative mesons stopped in matter are always captured (indicating a strong interaction with nuclei). Here presumably Eq. (8) describes the process, and the expected happens; the capturing nucleus explodes in a shower of particles known as a *star* due to its characteristic appearance in the photographic emulsions used for this high altitude research. It was evident that the mesons which produced these events (stars) belonged to a different species than those observed at sea level, and they were, consequently, designated as σ-mesons.

The existence of a new type of meson was directly established by Powell and his coworkers[21] in 1947. They measured the masses of the particles which produced tracks in photographic plates exposed at high altitudes and found them to be about 300 electron masses. These particles, named π-mesons, occurred both positively and negatively charged and it was observed that when stopped in the emulsion, the negative π-mesons always ended in stars. (Hence, the previously observed σ-mesons are negative π-mesons.) To avoid confusion with π-mesons, the previously known mesons occurring at sea level were called μ-mesons. These investigators found examples of the decay of

[20] J. Daudin, *Compt. Rend.* **218**, 830 (1944); W. E. Hazen, *Physical Rev.* **63**, 213 (1943) and **65**, 67 (1944); W. M. Powell, *Physical Rev.* **69**, 385 (1946) and D. H. Perkins, *Nature* **159**, 126 (1947) and **160**, 707 (1947). See also Ref. 1.

[21] Lattes, Muirhead, Occhialini, and Powell, *Nature* **159**, 93, 186, 694 (1947) and **160**, 453, 486 (1947).

π-mesons into lighter particles which were tentatively identified as μ-mesons. Since the decay must involve at least two products, the unobserved one was presumed to be neutral.

New evidence also rendered the decay of the ordinary μ-meson incapable of explanation by means of the assumptions of the old theory. It was assumed previously that a negative (positive) μ-meson decayed into an electron (positron) and a neutrino. The laws of conservation of momentum and energy then require that a meson at rest which decays will always produce an electron or positron of unique energy (about 50 Mev). However, the accurate results of more recent investigators[22] do not bear out the existence of this unique electron energy. Instead, the energies of the decay electrons have a continuous distribution which implies a decay of the meson into at least *three* particles. Only one of these three particles, the electron, is charged and hence observable. Experimental methods cannot yet inform us as to the nature of the neutral decay products.

As soon as the 380-Mev, 184-in. cyclotron at Berkeley began operation, experimental meson research gained an invaluable tool. For the first time, controllable energies of cosmic-ray magnitude were available and the artificial production of mesons could be attempted.

Gardner and Lattes[23] first succeeded in producing mesons by the collision of 380-Mev alpha-particles with target nuclei. These proved to be π-mesons of mass about 300 electron masses, since, when stopped in photographic plates, the negative particles are always captured by nuclei with the production of stars. These π-mesons were unstable, just as were the high altitude π-mesons, and presumably one of their decay products was a μ-meson. Richardson[24] was able to measure their lifetime and found it to be of the order of 10^{-8} sec.

V. Attempts to Obtain a Consistent Theory

An attempt can now be made to build a consistent scheme which will at least qualitatively account for the above facts. The so-called, two-meson hypotheses were first proposed by Marshak and Bethe[25] and independently by S. Sakata and T. Inowe.[26] Neither of these proposals agreed in detail with the one now generally accepted.[27]

In outline, the present scheme assumes that the heavy, or π-mesons, interact strongly with nucleons and are strongly produced in the collisions of the high energy primary protons with the nuclei in the upper layer of the earth's atmosphere. These mesons are the ones responsible for nuclear forces due to this strong interaction. Their mass (300 electron masses) agrees better with that predicted theoretically by Yukawa than does the mass of the μ-meson. The stars invariably observed at high altitude when negative π-mesons are stopped in matter also confirm this strong interaction, i.e., in spite of their short lifetime of 10^{-8} sec the negative π-mesons react so strongly with matter that capture can always successfully compete with spontaneous decay. The absence of all but a few π-mesons at sea level (hence the early failure to observe them) and the predominant presence of μ-mesons is accounted for by this spontaneous decay of π's into μ's with this short lifetime.

The μ-mesons are assumed to have a very small interaction with nucleons; this explains the absence of scattering and the dependence of the capture of negative μ-mesons on the atomic number of the absorber. Since the nuclear interaction of μ-mesons is so small, the spontaneous decay (lifetime 2×10^{-6} sec) can successfully compete with the nuclear capture in light elements, while the nuclear capture successfully competes with the spontaneous decay in heavy elements. The continuous energy spectrum of the observable electrons from μ-decay is explained by the decay of the μ-meson into three particles.

The explanation of the presence of stars following π-capture and their absence following μ-capture is to be sought in the spin assignments of the two types of mesons and in the detailed

[22] Anderson, Adams, Lloyd, and Rau, *Physical Rev.* **72**, 724 (1947); Adams, Anderson, Lloyd, Rau, and Saxena, *Rev. Mod. Physics* **20**, 334 (1948); and V. T. Sigurgeirson and K. A. Yamakawa, *Rev. Mod. Physics* **21**, 144 (1949).

[23] E. Gardner and C. M. G. Lattes, *Science* **107**, 270 (1948); Burfening, Gardner, and Lattes, *Physical Rev.* **75**, 382 (1949).

[24] J. R. Richardson, *Physical Rev.* **74**, 1720 (1948).

[25] R. E. Marshak and H. A. Bethe, *Physical Rev.* **72**, 506 (1947).

[26] S. Sakata and T. Inowe, *Progr. Theor. Physics* **1**, 143 (1946).

[27] R. E. Marshak, *Physical Rev.* **75**, 700 (1949); Y. Tanikawa, *Progr. Theor. Physics* **2**, 220 (1947); S. Hayakawa, *Progr. Theor. Physics* **2**, 200 (1947); and Taketani, Nakamura, Ono, and Sasaki, *Physical Rev.* **76**, 60 (1949).

reactions which are assumed to describe the above processes. A μ-meson can be assumed to decay into an electron and two neutrinos. Although the possibility that the neutral decay products consist in part of neutral mesons cannot be excluded, the above assumption is made for simplicity. Tiomno and Wheeler[28] have calculated the electron spectra for μ-meson decay under the assumptions of several different masses, both zero (neutrinos) and finite, for the neutral decay products. The experiments are not yet precise enough to decide which one best explains the facts. If the neutral decay products are two neutrinos, the μ-meson decays into three particles of spin $\frac{1}{2}$ and hence by the rule of vector addition must itself have spin $\frac{1}{2}$

$$\begin{array}{ccccccc} \mu^{\pm} & \rightarrow & e^{\pm} & + & \nu & + & \nu \\ \frac{1}{2} & & \frac{1}{2} & & \frac{1}{2} & & \frac{1}{2}, \end{array} \qquad (9)$$

where the spins of the particles are explicitly indicated. Spin $\frac{3}{2}$ is also consistent with Eq. (9), but is excluded for simplicity.

It can be mentioned here that spin 1 (or 0) could be consistently assigned to the μ-meson, still preserving the three-particle decay, if one of the neutral decay products in Eq. (9) were a neutral meson of integral spin (0 or 1), or if the μ-meson decayed into an electron, a neutrino, and a photon

$$\begin{array}{ccccccc} \mu^{\pm} & \rightarrow & e^{\pm} & + & \nu & + & \text{photon} \\ (0 \text{ or } 1) & & \frac{1}{2} & & \frac{1}{2} & & 0 \quad . \end{array} \qquad (10)$$

However, neither of these schemes would fit the facts of π-μ-decay (see below) as well as Eq. (9). In addition, the decay of a μ-meson of integral spin with the emission of a photon has been shown theoretically by Feer[29] to be very improbable. Several observers[30] searched for photons associated with μ-meson decay and detected none.

Some of the properties of the π-meson may now be inferred from knowledge of the π-μ-decay and the above assumptions about μ-mesons. It is quite probable from the data that a π-meson at rest decays into a μ-meson of definite energy. This implies a two-particle decay; the second particle is uncharged since it has not been detected and since the charges of the π- and μ-mesons provide the necessary conservation of charge in the process. From a knowledge of the masses of the π- and μ-mesons and the unique energy of the emitted μ-meson, the mass of the neutral decay product can be calculated by use of the laws of conservation of energy and momentum. Brode[31] gives the masses of the π- and μ-mesons as 283 ± 7 and 215 ± 5 electron masses, respectively. These values make the mass of the neutral product most probably zero (i.e., a neutrino). If the μ-meson has spin $\frac{1}{2}$ as assumed above, this implies that the π-meson has integral spin (0 or 1) so that the reaction of π-μ-decay with spin assignment becomes

$$\begin{array}{ccccc} \pi^{\pm} & \rightarrow & \mu^{\pm} & + & \nu \\ (0 \text{ or } 1) & & \frac{1}{2} & & \frac{1}{2}. \end{array} \qquad (11)$$

This integral spin assignment to the π-mesons together with their presumed strong nuclear interaction allows them to be postulated as the agency of nuclear forces and allows an explanation of star production by the nuclear capture of negative π-mesons. The mechanism of nuclear charge-exchange forces is similar to that given by Eq. (5) on the basis of the old incorrect theory

$$\begin{array}{ccccccccccc} N & + & P & \rightarrow & \pi^{-} & + & P' & + & P & \rightarrow & P' + N' \\ \frac{1}{2} & & \frac{1}{2} & & (0 \text{ or } 1) & & \frac{1}{2} & & \frac{1}{2} & & \frac{1}{2} \quad \frac{1}{2} \end{array} \qquad (12)$$

and

$$\begin{array}{ccccccccccc} N & + & P & \rightarrow & N & + & N' & + & \pi^{+} & \rightarrow & P' + N' \\ \frac{1}{2} & & \frac{1}{2} & & \frac{1}{2} & & \frac{1}{2} & & (0 \text{ or } 1) & & \frac{1}{2} \quad \frac{1}{2}. \end{array}$$

The capture of negative π-mesons presumably entails the absorption of the π-meson by a proton in the nucleus

$$\begin{array}{ccccc} P & + & \pi^{-} & \rightarrow & N \\ \frac{1}{2} & & (0 \text{ or } 1) & & \frac{1}{2}. \end{array} \qquad (13)$$

After this absorption the neutron has an excitation energy at least equal to the mass energy of the π-meson, about 140 Mev. This excitation energy produces an evaporation of the nucleus with the observation of a nuclear explosion or star.

To set up a reasonable reaction with a consistent spin assignment describing the capture of a negative μ-meson is a more difficult task

[28] J. Tiomno and J. A. Wheeler, *Rev. Mod. Physics* **21**, 144 (1949).
[29] D. B. Feer, *Physical Rev.* **75**, 731 (1949).
[30] R. D. Sard and E. J. Althaus, *Physical Rev.* **73**, 1251 (1948); O. Piccioni, *Physical Rev.* **74**, 1236 (1948); and E. P. Hincks and B. Pontecorvo, *Physical Rev.* **73**, 257 (1948).
[31] R. B. Brode, *Rev. Mod. Physics* **21**, 37 (1949).

since the absence of any star or subsequently emitted particle to dissipate the mesons rest energy (10^7 Mev) must be accounted for. The assignment of spin $\frac{1}{2}$ to the μ-meson makes a reaction like Eq. (13) impossible. This is fortunate for it enables another particle to be introduced to *conserve spin* and hence this particle can be postulated to be unobservable (i.e., a neutrino or neutral meson) and can serve to carry off this excess energy. A process that can be envisaged is a two-step one in which the μ-meson first virtually changes into a π-meson and a neutrino; the π-meson is then absorbed by a proton in the nucleus while the unobservable neutrino carries off most of the rest energy of the absorbed μ-meson. This reaction with a consistent spin assignment would be

$$\begin{array}{cccccccccc} \mu^- & + & P & \xrightarrow{\text{(virtually)}} & \pi^- & + & \nu & + & P & \rightarrow N+\nu \\ \frac{1}{2} & & \frac{1}{2} & & (0\text{ or }1) & & \frac{1}{2} & & \frac{1}{2} & \frac{1}{2}\quad\frac{1}{2}. \end{array} \qquad (14)$$

Rosenbluth[32] has performed calculations with models similar to this and finds that although the neutrino carries off most of the energy, the remaining neutron will be left with an excitation energy of 10 to 15 Mev. This energy is not sufficient to disrupt the nucleus and produce a star, but it is sufficient to allow the escape of the neutron, since, on the average, neutrons are bound in nuclei with energies of about 8 Mev.

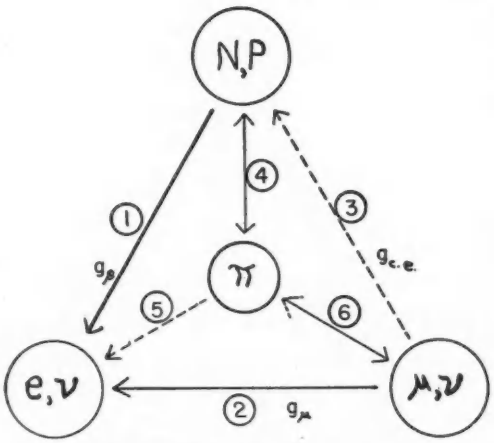


FIG. 2. The interactions between the various elementary particles.

[32] M. N. Rosenbluth, *Physical Rev.* **75**, 532 (1949); Lee, Rosenbluth, and Young, *Physical Rev.* **75**, 905 (1949).

Both Sard[33] and McClure and Groetzinger[34] report the emission of neutrons from absorbers in which negative μ-mesons have been observed to stop, thus lending strength to the argument.

Tiomno and Wheeler[35] propose a reaction (charge-exchange reaction) alternate to Eq. (14); the nucleon ($P \rightarrow N$) is imagined to be coupled to, and to interact with, both the μ-meson and the neutrino so that the process may occur in one step without the virtual intermediate π-step

$$\begin{array}{cccccc} \mu^- & + & P & \rightarrow & N & + \nu \\ \frac{1}{2} & & \frac{1}{2} & & \frac{1}{2} & \frac{1}{2}. \end{array} \qquad (15)$$

It should be pointed out that a neutral meson cannot be excluded from constituting the neutral decay particle in Eqs. (11) and (14) or (15). The neutrino is chosen both for simplicity and because it fits the observed masses of π- and μ-mesons better.

The various interactions between the elementary particles can be succinctly represented by means of a diagram due mainly to Yukawa.[36] The solid lines represent processes and interactions which are well verified: ② is μ-meson decay; ⑥ is π-μ-decay; and ④ represents the interaction responsible for nuclear forces and π-meson production. It will be noted that ①, a direct interaction between nucleons on one hand and electrons-neutrinos on the other, corresponding to the reaction Eq. (6), has to be reintroduced in order to account for the β-decay of nuclei. Marshak[37] has suggested the interaction ⑤ coupling the π-meson with electrons and neutrinos so that ① could be eliminated and β-decay could be accounted for by

$$\begin{array}{cccccccc} N & \rightarrow P + & \pi^- & \rightarrow P + & e^- & + \nu \\ \frac{1}{2} & \quad\frac{1}{2} & (0\text{ or }1) & \quad\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{array} \qquad (16)$$

which is similar to Eq. (7) of the old theory. However, no such decay of π-mesons into electrons has been observed.

The capture of μ-mesons by nuclei, Eq. (14), involves two steps and the two interactions ④

[33] Sard, Ittner, Conforto, and Crouch, *Physical Rev.* **74**, 97 (1948); R. D. Sard, *Rev. Mod. Physics* **21**, 35 (1949).

[34] G. W. McClure and G. Groetzinger, *Physical Rev.* **75**, 340 (1949).

[35] J. Tiomno and J. A. Wheeler, *Rev. Mod. Physics* **21**, 153 (1949).

[36] H. Yukawa, *Mass. Inst. Tech. Coll. Lect.* (December 9, 1948). See also Ref. 35.

[37] R. E. Marshak and H. A. Bethe, *Physical Rev.* **72**, 506 (1947).

and ⑥. Wheeler's alternate direct interaction, Eq. (15), is represented by ③. The coupling constants g_β, g_μ, and $g_{c.c.}$ are analogous to e, the electronic charge, which is the coupling constant between the electron and the electromagnetic field. These determine the strength of the corresponding interactions. They can be inferred from the rate of β-decay, the rate of μ-meson decay, and the capture probability of μ-mesons by nuclei, respectively. Wheeler and others[38] point out that in numerical magnitude

$$g_\beta \sim g_\mu \sim g_{c.c.} \qquad (17)$$

within experimental and calculational error. This may be accidental, but since these three processes all involve neutrinos, the above authors suggest that the approximate equality may indicate some fundamental connection between all processes involving neutrinos that so far has escaped general notice.

VI. Conclusion

The above scheme correlates and accounts for, at least qualitatively, most of the present information about mesons and their interactions with the various other elementary particles. It should be emphasized that the scheme does not yet yield quantitatively exact results. This is particularly true about the status of nuclear forces in regard to which current experimentation with the new high energy machines is yielding data about nuclear interactions which cannot be accounted for by means of the present meson theories.

Other findings are beginning to cast doubt upon the correctness of this picture in all its details. One of the most striking and important of these is the recent identification of a new meson (called κ- or τ-meson) of mass in the region 800-1000 electron masses by Leprince-Ringuet and others.[39] There are very little data yet available on this subject so it is too soon to say how this meson could fit into a consistent scheme. However, Tanikawa[40] has made some tentative suggestions as to how such a very heavy meson could be fitted into the scheme.

The scheme outlined in Section V will most probably be changed and modified in the light of present and future experimental discoveries. However, its qualitative explanations of many of the presently known and well-substantiated phenomena indicate that many of its features must be in the main correct.

[38] J. Tiomno and J. A. Wheeler, *Rev. Mod. Physics* **21**, 153 (1949); Lee, Rosenbluth, and Young, *Physical Rev.* **75**, 905 (1949).

[39] L. Leprince-Ringuet and M. Lheritier, *Compt. Rend.* **219**, 618 (1944); L. Leprince-Ringuet, *Rev. Mod. Physics* **21**, 42 (1949); G. D. Rochester and C. C. Butler, *Nature* **20**, 855 (1947); and R. Brode, *Rev. Mod. Physics* **21**, 37 (1949).

[40] Y. Tanikawa, *Progr. Theor. Physics* **3**, 314, 315 (1948).

Opportunities for Graduate Study in Physics

The Editor has a supply of reprints of an announcement bearing the above title, published in this Journal in the February, 1949 issue. As long as the supply lasts he will be glad to send copies free of charge to interested teachers and students. Requests should be addressed to Thomas H. Osgood, *Editor*, American Journal of Physics, Michigan State College, East Lansing, Michigan.

NOTES AND DISCUSSION

The Capacitor-Resistor Circuit

LIONEL FLEISCHMANN
1217 Jefferson Avenue, Toledo 2, Ohio

ALL of the many textbooks consulted by me give the formula for the voltage transient v_c (volts) across the capacity C (farads) in series with a resistance R (ohms) upon application of a constant emf E (volts) as

$$v_c = E(1-e^{-t/RC}), \tag{1}$$

where t is the time in seconds and $v_c=0$ for $t=0$.[1] The current i (amp) is given by

$$C(dv_c/dt) = (E/R)e^{-t/RC}, \tag{2}$$

which for $t=0$ leads to $i=(E/R)$.

Also, all the textbooks which I have consulted, with the exception of one, point out that the current at the moment of closing has the finite value (E/R), and one by a prominent author goes so far as to state: "The important thing to note, however, is that with capacity the current is always a maximum at $t=0$. The reason for this is clear. If the condenser is empty, the electrical pressure for discharge is then the greatest. Hence, capacity stimulates flow of current."

The one textbook[2] mentioned as an exception shows an oscillogram with the current (as it must be) zero at the time of closing, rising rapidly and then falling off. The accompanying text states: "In Fig. 425 is an oscillogram of the current in the circuit of Fig. 423. Any circuit has some inductance and the inductance prevents the instantaneous change of current." It goes on to say, after some remarks referring to the bifilar type of oscillograph used, "however, during a very short interval of time following the closing of the circuit, the current values as determined from the oscillogram will check closely those computed from Eq. (2)."

It is the object of the following discussion to show that a derivation of the condenser equations which are in accord with the physical conditions can be easily obtained. But before doing this it should be pointed out that the method for arriving at Eqs. (1) and (2) is wrong in principle, because it does not take into account the impossibility of the current being zero and finite at the same time.

The reason for accepting this impossibility seems to have its origin in the apparent similarity of the capacity resistance circuit and the self-induction-resistance circuit. With a self-induction of L (henries) in series with a resistance of R (ohms), upon closing the circuit, the equation for the current transient i (amp) produced by constant emf E (volts) is

$$L(di/dt)+Ri=E. \tag{3}$$

For the determination of the constant of integration we have the condition $i=0$ for $t=0$, which implies also that for $t=0$, $L(di/dt)=E$. But there are no physical conditions contradicting this and a linear differential equation of the first order is, therefore, sufficient. With the same circuit constants as used in Eq. (1), the transient voltage across the capacitor in the capacity-resistance circuit is given by $v_c+RC(dv_c/dt)=E$, with the two conditions $v_c=0$ for $t=0$ and the current $C(dv_c/dt)=0$ for $t=0$. It is customary to neglect the current condition, and the solution for v_c is then

$$v_c = E(1-e^{-1/CRt}).$$

Trying to use the condition of zero current at time $t=0$ for the determination of the constant would lead to constant$=0$, which is useless. The answer to this dilemma is that in order to obtain a solution with two arbitrary constants we must have a differential equation of the second order. The physical condition corresponding to the mathematical demand is found by introduction of a self-induction in series with both the capacity and resistance. The question could be raised why is it admissible to neglect the capacity in a self-induction-resistance circuit, and the answer is that it is physically possible to realize such a circuit without any series capacity. While in the case of the capacity-resistance circuit, no matter how short we make the leads and how near we place them, we shall always have finite self-induction, and therefore, with L (henry) self-induction, the differential equation for the transient condenser voltage is

$$LC\frac{d^2v_c}{dt^2}+RC\frac{dv_c}{dt}+v_c=E$$

with the two limiting conditions $v_c=0$ for $t=0$, and $C(dv_c/dt)=0$ for $t=0$. The solution is

$$v_c=(E/2f)\{(R/2L-f)\exp(-Rt/2L-ft) -(R/2L+f)\exp(-Rt/2L+ft)\}+E, \tag{4}$$

where f is written for $\{(R/2L)^2-(1/LC)\}^{\frac{1}{2}}$.

We will now introduce the condition $(R/2L)^2 \gg (1/LC)$ and L very small; then we can substitute for f the expression $[(R/2L)-(1/CR)]$ and the current will be

$$i=\frac{E}{[R-(2L/CR)]}[\exp(-t/CR)-\exp(-Rt/L+t/CR)]. \tag{5}$$

The current will reach a maximum for the time t_0 obtained from the equation $(di/dt)=0$

$$t_0=\frac{2.303\log[(CR^2/L)-1]}{[(R/L)-(2/CR)]}. \tag{6}$$

To give an idea of the difference of maximum currents when using Eq. (5) as compared with Eq. (2), we will assume $L=2\times10^{-7}$ henry (which is the value for two meters of a pair of bare, cylindrical conductors placed side-by-side), $C=10^{-6}$ farad and $R=2$ ohm. Inserting these values in Eq. (6) we obtain $t_0=3.3\times10^{-7}$ second. Introducing 3.3×10^{-7} second in Eq. (5) furnishes $i_0=E\times0.448$ amp against $E\times0.5$ amp from Eq. (2), or about 10 percent difference. For any time t sec, the difference between the fictitious current and the actual current is given by

$$E\frac{Re^{(-R/L+1/CR)t}-(2L/CR)e^{-t/CR}}{(R^2-2L/C)}. \tag{7}$$

Both currents will have the same value for time

$$t_1(\text{sec}) = 2.303 \frac{\log(2L/CR^2)}{(-R/L+2/CR)}. \quad (8)$$

Inserting values as above we find $t_1 = 2.56\times10^{-7}$ sec and the current $E\times0.439$ amp. Equation (5) satisfies the condition that $i=0$ for $t=0$, and the current after passing through a maximum at t_0 given by Eq. (6) tends toward zero for $t\to\infty$.

For the discharge of a capacitor charged to a voltage V_c we have the equation

$$LC\frac{d^2v_c}{dt^2}+RC\frac{dv_c}{dt}+v_c=0$$

with the conditions $v_c = V_c$ for $t=0$, and $i = C(dv_c/dt) = 0$ for $t=0$. As nothing new in principle is added by giving their solutions, these are omitted. But as with the charging process some time elapses after closing the circuit before the current reaches its maximum.

In conclusion, I would say that the textbooks in treating this subject should mention that the simplified circuit without self-induction is a mathematical fiction and that the current, no matter how near its maximum comes to be equal to (E/R), can never attain this value at the moment of closing the circuit.

[1] We only consider the case of $v_c=0$ for $t=0$, but an initial charge of the capacity does not affect the final conclusions.
[2] Mueller, *Introduction to electrical engineering* (McGraw-Hill, ed. 2, 1948), p. 502.

AMERICAN JOURNAL OF PHYSICS VOLUME 18, NUMBER 1 JANUARY, 1950

LETTERS TO THE EDITOR

Significant Figures in the General Physics Course

THE role of significant figures in the handling of numerical data is always somewhat confusing to the beginning student and is usually taught in an introductory physics course by means of "rules of thumb."

These rules of thumb, while quite useful and correct, are generally introduced in an arbitrary manner. It has, however, been felt here that a straightforward illustration of the concepts involved and the required arithmetical manipulations is desirable when this subject is introduced in the beginning lectures of the first-year course. (It might even be a good idea if reintroduced into the first few lectures of the second-year course!)

To begin with, the measurement of length is discussed. After it is agreed that an arbitrary standard for comparison exists, the choice of subdivisions of the fundamental unit is considered. Thus, we might take a meter stick and divide it into 10 equal parts, or 100, or 1000 equal parts. We find that if we must measure a length which is not an integral number of a chosen group of subdivisions, we are faced with the problem of estimating the left-over fraction of the subdivision, and we thus introduce the concept of *region of uncertainty* due to the observer alone, ignoring here any other sources of error. It becomes quite clear that even a well-trained observer will tend to estimate the value of this fraction differently with subsequent readings. It is also quite clear, that if one increases the number of subdivisions of the fundamental unit (within reason!), that the magnitude of this region of uncertainty will decrease.

Thus, if we assume, for the sake of argument, that $\pm\frac{1}{10}$ of any given subdivision will be the region of uncertainty, a meter stick marked off in 10-cm subdivisions will have a $\pm$1-cm region of uncertainty; 1-cm subdivisions will give the meter stick a $\pm$1-mm region of uncertainty, and so on.

As an example, we may apply this to the problem of determining both the perimeter and the area of a rectangle. Let us assume that the rectangle has sides which have nominal values of A_0 and B_0 cm, and that the rod used to measure the length A_0 cm has a region of uncertainty of $\pm a$ cm, while the rod used to measure the length B_0 cm has a region of uncertainty of $\pm b$ cm.

The measurement of the two lengths can then be written: $(A_0\pm a)$ cm and $(B_0\pm b)$ cm. By factoring we write $A_0(1\pm a/A_0)$ cm, etc., where $\pm(a/A_0)\times100$ is the percentage error in the measurement of A_0. A distinction is obviously made between *region of uncertainty* which is a quantity of the dimensions of the measured quantity, and *percent error* or *percent uncertainty* which is a pure number.

To determine the perimeter of the rectangle, we add the two measurements and multiply by two.

$$P = 2[(A_0\pm a)+(B_0\pm b)] = 2[(A_0+B_0)\pm(a+b)],$$

with the result that we simply add the regions of uncertainty. If we factor, we have

$$P = 2(A_0+B_0)\left[1\pm\frac{(a+b)}{(A_0+B_0)}\right], \quad (1)$$

and the term $\pm[(a+b)/(A_0+B_0)]\times100$ is the percentage uncertainty in the determination of the perimeter. Before considering this further, let us obtain the area S of the rectangle:

$$S = (A_0\pm a)(B_0\pm b) = A_0B_0(1\pm a/A_0)(1\pm b/B_0).$$

Carrying the multiplication through, and observing that both $|a/A_0|$ and $|b/B_0| \ll 1$, we get

$$S = A_0B_0\left[1\pm\left(\frac{a}{A_0}+\frac{b}{B_0}\right)\right], \quad (2)$$

where $\pm[(a/A_0)+(b/B_0)]\times100$ is the percentage uncertainty in the determination of the area. Equations (1) and (2) can now be used to discuss how one handles

numbers which are the results of measurement. Thus the addition or subtraction of such numbers is governed by Eq. (1) and their multiplication or division by Eq. (2).

To summarize, we may deduce these *rules* for the manipulation of numbers which are results of physical measurements:

1. (a) When adding (subtracting) such numbers, the region of uncertainty of the result is the sum of the regions of uncertainty of all the numbers involved; (b) The percentage uncertainty of the result, however, is of the order of magnitude of the largest percentage uncertainty present (see Eq. (1));
2. When multiplying (dividing) such numbers, the percentage uncertainty of the result is the *sum* of the percents of uncertainty of all the numbers involved.

We may note, as a corollary to Eq. (2), that if we have numerical results of measurements, A, B, C, $\cdots$, with appropriate regions of uncertainty ΔA, ΔB, ΔC, $\cdots$, and if S is the nominal result of multiplying A, B, C, $\cdots$,

$$S=A\times B\times C\times\cdots. \tag{3}$$

Then, if ΔS is the region of uncertainty in S, the percentage uncertainty of the result is

$$\pm\frac{\Delta S}{S}=\pm\left(\frac{\Delta A}{A}+\frac{\Delta B}{B}+\frac{\Delta C}{C}+\cdots\right). \tag{4}$$

This is immediately obtained by taking the logarithmic derivative of Eq. (3).

The geometric interpretation of Eqs. (2) and (3) is obvious, but is useful in aiding the student in their comprehension.

The power of 10 notation should be used consistently since it removes any doubts as to which zeros are to be included in the number as "significant" figures and which zeros are present to place properly the decimal point.

HAROLD FLEISHER
LEONARD O. OLSEN

Case Institute of Technology
Cleveland, Ohio

Further Comments on the Concept of Pressure

IN a previous article,[1] the author presented a concept of pressure which is held by many experimental physicists. The definition of pressure presented was that pressure on a large surface where the force is uniform means the force that is exerted on a unit area, and the pressure has the direction of that force. Using this definition, pressure is a vector quantity. A concrete example may be taken from an experiment. On a horizontal surface facing upward at some fixed position in a liquid, there is a pressure p_1 down on that surface, which pressure can be measured. Whenever a pressure on a surface is measured, that pressure, which is the force on unit area, has direction. Some may claim that what is measured is the force on unit area, and they may not want to use the term pressure for this measured force on unit area. These people have a right to define pressure as they desire; but many experimental physicists do define pressure as that force on unit area.

The following remarks apply to Summers' four points of disagreement[2] with the above-stated definition of pressure.

1. Using the above definition, which is more completely given in the original article[1] by the author, pressure is a vector quantity. The author completely agrees with Summers in his statement, "It would certainly be undesirable to have fundamentally different concepts of pressure in elementary and in advanced work." The author would add that it is undesirable that the term pressure be a vector as used by an experimental physicist, and be a scalar as used by a mathematical physicist. The whole purpose of these articles is to come to an agreement, if possible, as to the definition of pressure. If that is not possible, then an author should define the way in which the term is being used.

2. According to Summers,[3] pressure is purely a ratio number with units, as the gravitational constant G is a ratio number with units. Using pressure in this way, Summers says that $\int pdv$ = work. Mathematically he comes out with the correct magnitude of the work, and his rule of units gives the correct units for the work. But to an experimental physicist who is interested in concepts of nature, this is not satisfactory. The experimental physicist has the mental concept of a force f_1 being exerted on a unit area, and in a simple case for the sake of clarity of the point under discussion, that unit area is pushed through a distance x. Thus, the external work done in a reversible process equals $\mathbf{f}\cdot\mathbf{x}$ or $\int\mathbf{f}\cdot d\mathbf{x}$. This has a physical meaning with a concept of what is happening in nature, and the terms force, pressure, distance, and work are defined to give meaning to what is taking place. Now it is true that for ease of calculating in many cases $\int f dx = \int p dv$ is used. But to an experimental physicist the concept is not changed

$$\int p dv=\int\frac{f}{\text{area}}dx\cdot\text{area}.$$

3. Experimentation shows the physicist that if a surface is placed within a fluid at rest, there is a force normal to that surface, and the force on unit area the physicist calls pressure. If the center of that surface is kept fixed with respect to the liquid, the pressure on that surface is of the same amount for each position of the surface, but the direction of that pressure changes as the direction of the surface changes. The physicist who is interested in physical concepts interprets the above experimental fact by saying that a surface of negligible thickness has a pressure of the same amount exerted on the two sides of that surface, irrespective of the position of the surface. Or in other words, the pressure at any place in the liquid is of the same amount and in all directions. This reasoning cannot be extended to a mathematical point within the liquid. Nothing is experimentally known of the pressure within an atom, a nucleus or an electron. Terms are defined and used by physicists to discuss experiments. If these terms are used in theory, usually their definitions are not changed. In the case of a microscopic body suspended in a liquid

at rest, we know that at a given instant the pressure is not the same in all directions on that body, for it can be observed to exhibit Brownian movements. Hence, when the statement is made that the pressure is the same in all directions at a given place in a liquid, a definite set of conditions as indicated above is assumed to exist.

All experimental evidence indicates that any body can push out in all directions at the same time, if there are bodies pushing on that body in all directions. The truth of this is readily detected if one gets in a crowd, where people are pushing on one from all directions. One will find that he is pushing back in each direction opposite to that in which he is being pushed.

4. It may not seem logical to some people that pressure and force are exactly similar concepts, and that the only difference in the terms is that pressure is the force on a unit area. But, to many a physicist, a word which expresses an idea may be defined as expressing that idea, and to that physicist, that definition is logical.

Hence to many physicists pressure is a force, but it is a force on a unit area. That area is given in what are called the units. Hence a pressure of 10 g per sq. cm on a large surface where the pressure is uniform, indicates the concept that a force of 10 g is exerted in a definite direction on 1 sq. cm. To these physicists, this pressure does not mean a ratio with units of g/sq. cm. It would indeed be desirable to come to an agreement between these two definitions of pressure. Does the term pressure as used in physics mean a ratio with units, or does it mean a force on a unit area, the pressure having the direction of the force?

VOLA P. BARTON

Goucher College
Baltimore, Maryland

[1] Vola P. Barton, *Am. J. Physics* **17**, 318–319 (1949).
[2] R. D. Summers, *Am. J. Physics* **17**, 319 (1949).
[3] R. D. Summers, *Am. J. Physics* **14**, 311–313 (1946).

Apparatus of Historical Interest

PHYSICS has been taught at Washington and Jefferson College since 1802. In 1852 Samuel Jones, a graduate of Trinity College, Dublin, became Professor of Physics. Quite a bit of the apparatus used and made by Jones remains today. The professor appointed in 1885 was James F. Ray who had studied under Kirchhoff in Berlin. He started student laboratory work. The department at present has collected some of the historical equipment and has placed it in display cases in Thistle Laboratory. Does anyone know the approximate ages of the more interesting pieces listed below?

Model steam engine: Manufactured by L. E. Knott Apparatus Company, Boston. Model shows cross-section view of cylinder showing movement of piston and valves. Over-all length is 23 cm; height, 14 cm; and flywheel, 9.5 cm in diameter.

Four-prism spectrometer: Manufactured by John Browning, London. Length of telescope is 36 cm; collimator, 33 cm; height from base to scale is 21 cm, to prism table, 24 cm. Prism angles are 50 degrees. Telescope has Gaussian eyepiece; Steinheil mounting.

Spectroscope: Made in Switzerland. Nameplate—Societé Genevoise Pour La Construction d'Instruments de Physique, Genève. parts: Illuminated scale; telescope, 17.8 cm long; collimator, 16.5 cm long; over-all height, 27.3 cm.

Electroplating apparatus: Manufactured by W. G. Pye and Company, Cambridge, England. Consists of small copper dish 7 cm in diameter, resting on a copper base used as one terminal; and a copper plug suspended inside the cup. Over-all height, 15 cm.

Galvanometer: Manufactured by Elliott Brothers, London. Clock-shaped wooden case; 18 cm in height, 6 cm in width. Scale reads 70–0–70.

Ammeter: Manufactured by Waterhouse Electric & Manufacturing Company, Hartford, Connecticut. Direct reading ammeter, range 0 to 45 amp. Rectangular wooden case 10.5 cm×11 cm×15 cm; scale on vertical face; one terminal on each side.

The department is also trying to find the age of a 6-inch refracting telescope made in New York by H. Fitz, at least 60 years ago.

RAYMOND M. BELL

Washington and Jefferson College
Washington, Pennsylvania

AMERICAN JOURNAL OF PHYSICS VOLUME 18, NUMBER 1 JANUARY, 1950

ANNOUNCEMENTS AND NEWS

Book Review

Introduction to Statistical Mechanics. RONALD W. GURNEY. Pp. 268, Figs. 59. McGraw-Hill Book Company, Inc., New York, 1949. Price $5.00.

Professor Gurney has written a very interesting and readable book on statistical mechanics which should prove valuable to graduate students of physics and chemistry. In the main, he has tried to present the fundamental methods of the subject and to illustrate these through a relatively few number of examples. Thus, this is a book which would allow the student to grasp the fundamental notions of statistical mechanics but which would not serve the purpose of a reference book. It contains little of

classical kinetic theory. For example, the transport phenomena, such as viscosity and heat conduction, are not treated nor is the concept of mean free path developed. One looks in vain in the index for such topics as the ergodic hypothesis, detailed balancing, or microcanonical ensemble. Clearly, the scope of the book is limited; but within the region which Professor Gurney has chosen, the presentation is both detailed and understandable.

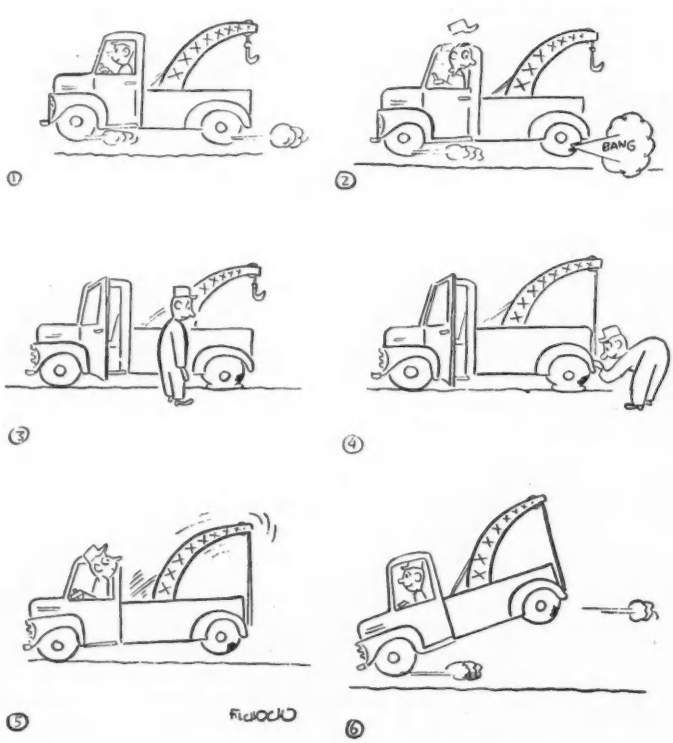
From the outset it is assumed that the energy states of the systems under consideration are quantized. It is only in the latter chapters that the statistical theory of classical systems is introduced and then, rather briefly.

In a rough way the book is divided into two parts, the first of which is concerned principally with a development and an understanding of the methods, while the second part deals with the applications. The first few chapters are devoted to the question—probably the most fundamental problem of statistical mechanics—of the existence of an equilibrium state and of its meaning. Professor Gurney unfolds the argument slowly and convincingly. The notion of temperature is carefully developed and this is followed by an introduction of the partition function and many of the properties of monatomic and diatomic gases are deduced. The final chapter of this first portion of the book presents the identification of the usual thermodynamic quantities such as entropy and free energy and shows the relationship between the thermodynamic laws and the results derived by statistical mechanics.

The applications are, on the whole, well chosen although sometimes rather specialized as in the case of the excellent discussion of solid and liquid solutions and of alpha- and gamma-iron. They include a treatment of nonideal gases, calculation of heat capacities, as well as a comparison between observed and theoretical chemical constants. The last chapter of the book describes the modifications introduced by the Fermi-Dirac and by the Bose-Einstein statistics and applies the former to the theory of free electrons in a metal.

Professor Gurney writes in his Preface that this is a book with a modest objective, namely to show that a subject, which is usually believed to be necessarily difficult and abstruse, can be presented in a form attractive to the experimentalist. Your reviewer feels the Professor Gurney has attained his objective and that he has written a clear and understandable account of that somewhat narrow region of statistical mechanics which he has chosen to expound.

DAVID M. DENNISON
University of Michigan



Action *vs.* reaction. (Reprinted from *Ethyl News*, June–July, 1949 by permission of the publisher and the artist.)